

# Machine learning-based prediction of cognitive trajectories among middle-aged and elderly Chinese adults

**Yafei Wu**

Xiamen University School of Public Health <https://orcid.org/0000-0003-3937-0263>

**Maoni Jia**

Xiamen University School of Public Health

**Chaoyi Xiang**

Xiamen University School of Public Health

**Shaowu Lin**

Xiamen University School of Public Health

**Zhongquan Jiang**

Xiamen University School of Public Health

**Ya Fang** (✉ [fangya@xmu.edu.cn](mailto:fangya@xmu.edu.cn))

Xiamen University School of Public Health <https://orcid.org/0000-0002-9895-3234>

---

## Research

**Keywords:** Cognitive trajectories, Machine learning, Growth mixture model, MMSE

**Posted Date:** May 18th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-524043/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

# Abstract

## Background

Cognition represents heterogeneous trajectories with increasing age. Early identification of such trend is essential for the prevention of high risk population related to cognition decline. This study aimed to explore the heterogeneity and determinants of cognitive trajectories, and construct prediction models for distinguishing cognitive trajectories among the middle-aged and elderly Chinese people at a community level.

## Methods

Data was retrospectively collected from the China Health and Retirement Longitudinal Study with a consecutive survey in 3 waves (2011, 2013, 2015). The Mini-mental State Examination (MMSE) scale was used for cognitive measurement. The heterogeneity of cognitive trajectories was identified through mixed growth model, and the determinants of cognitive trajectories were analyzed by logistic regression. Machine learning (ML) algorithms, namely logistic regression (LR), support vector machine (SVM), and random forest (RF), combined with recursive feature elimination were used to predict cognitive trajectories using epidemiological variables. Area under the receiver operating characteristic curve (AUROC) and brier score were used to assess discrimination and calibration, respectively.

## Results

Three cognitive trajectories were identified: “persistently-high” (PH, 83.3%) class with a high cognitive score in baseline and remained high subsequently, “medium-decrease” (MD, 6.6%) class with a medium cognitive score in baseline and a rapid decline subsequently, “low-increase” (LI, 10.1%) with the lowest baseline cognitive score and a rapid growth subsequently. For classification of PH vs. LI and MD vs. LI, the AUROCs of three ML methods were almost larger than 0.90, and brier scores were almost less than 0.10. For PH vs. MD, ML performed less well with the maximum AUROCs close to 0.68, while the calibration was still good. The initial MMSE score was the most significant predictors for distinguishing MD vs. LI and PH vs. LI, while age was the most important predictors in distinguishing MD vs. PH.

## Conclusions

Prediction of cognitive trajectories and identification of its key predictors are of great significance for understanding the heterogeneity of cognitive trajectories as well as interventions for cognitive decline.

## Background

Dementia is a degenerative disease of the central nervous system accompanied by cognitive decline. According to the World Alzheimer's Report, more than 50 million people worldwide were estimated to be

living with dementia in 2020. Moreover, the number will increase to 152 million by 2050 with rapid aging [1]. With frequently exposure to unhealthy lifestyles and environmental hazards, middle-aged people also faced the risk of cognitive decline for the rejuvenation of chronic diseases [2]. Until now, the etiology and pathogenesis of dementia are still unclear, and there is no effective cure in clinical practice [3]. About three quarters of dementia patients worldwide were not diagnosed in time, especially in low- and middle-income countries. It was estimated that the annual global losses related to dementia were \$818 billion, which was equivalent to the output of the 18th largest economy [1]. Therefore, it is crucial to identify the high risk individuals of cognitive decline and evaluate their cognitive trajectory for the timely prevention of dementia.

So far, most of the prediction studies related to cognition focused on clinical diagnosis [4]. Thus, data used in these studies were obtained from some specific clinical cohorts. Specifically, a lot of complex, costly, and invasive data including neuroimaging, cerebrospinal fluid, and genetic data were collected, which was not suitable for early screening in a more general population [5]. In addition, previous studies mainly considered cognition at a cross-sectional level [6], which could not reveal cognitive trajectories over time. So cognitive trajectory prediction based on longitudinal data may shed more light on the change of cognition.

For population trajectory analysis, traditional models are not suitable for its assumption of homogeneity of all samples. Growth mixture model (GMM), a combination of traditional growth model and latent class analysis, can identify the potential trajectories with a consideration of variation contained in different trajectories as well as population in the same trajectory, and has been gradually used in cognitive trajectory study [7]. In terms of cognition prediction, machine learning techniques are frequently applied. Machine learning (ML), which is a set of computational methods that can discover complex nonlinear relationships between inputs and outputs, has been widely used in disease prediction [8]. In medical practice, ML can process complex medical data and construct prediction models for decision-making.

This study focused on two issues: firstly, mixed growth model was used to explore the cognitive trajectories among the middle-aged and elderly people in China, and the determinants of cognitive trajectories were also explored. Second, machine learning methods, namely logistic regression, support vector machine, and random forest, were used to predict the types of cognitive trajectories with epidemiological variables, and the importance of predictors were assessed by random forest, so as to develop specific intervention strategies for high risk populations.

## Methods

### Data Source

This study retrospectively collected data from the China Health and Retirement Longitudinal Study (CHARLS, <http://charls.pku.edu.cn/>) with a consecutive survey in 2011, 2013, and 2015. CHARLS has been widely recognized and used in academia, as described elsewhere [9]. The three waves were used to identify cognitive trajectories classes for the middle-aged and elderly adults, and data in the first wave were used as predictors. Participants aged 45 and above with a consecutive survey during three waves were included. Participants with missing values in MMSE items were excluded. Finally, a total of 4,962 participants were

selected for model derivation and internal evaluation. Data used in study was approved by the biomedical ethics committee of Peking University, and all participants provided written informed consent.

## Measurement of variables

Predictors in this study included sociodemographics, behavior and lifestyle, baseline cognitive and depression scores, physical examination information. For sociodemographics, age, gender, marital status (not in married vs. in married), and education (illiteracy vs. non-illiterate) were selected. For behavior and lifestyle, smoking, drinking, sleep time, and social activities were considered. Smoking was obtained through the following question: "*Do you smoke now?*". Drinking was obtained through: "*Have you ever drunk in the past year?*". Both smoking and drinking were converted into dichotomous variables. Sleep time referred to the average sleep time in the past month, and social activities engagement were referred to participation in any social activities in the past month. Baseline cognitive scores, baseline depression scores were calculated through the Mini-Mental State Examination (MMSE) Scale and the Center for Epidemiologic Studies Depression Scale-10 (CESD-10) Scale [10–11]. Physical examination variables included waist circumference, chronic diseases (hypertension, diabetes, psychiatric problems, and memory-related diseases). Waist circumference referred to the circumference of the waist around the navel at the end of expiration when the subject was breathing calmly and is calculated in centimeters. Hypertension was defined as systolic blood pressure  $\geq 140$ mmHg and/or diastolic blood pressure  $\geq 90$ mmHg, or use of antihypertensive drugs. Diabetes referred to fasting blood glucose  $\geq 126$  mg/dL, HbA1c  $\geq 6.5\%$ , or current anti-diabetic treatments. Psychiatric problems, and memory-related diseases were collected through self-reported doctors' diagnosis. For missing values, multiple imputation was performed before constructing prediction models. We generated 5 sets of data and obtained the mean values for each index.

## Feature selection

Recursive feature elimination (RFE) is an ML method for feature selection that combines with several classifiers to eliminate redundant variables, thus identifying the most important factors for each classifier [12]. In order to select the best combination of predictors, a 10-fold cross validation was combined with RFE, that was, RFE was performed on each subset of input data, and validation error of all subsets was calculated. Finally, the subset with smallest error was selected as the optimal combination.

## Cognitive trajectories of middle-aged and elderly adults

Firstly, we performed a multiple linear regression to obtain the predicted cognitive score of each person adjusting for age, gender and education. Next, the adjusted  $Z$  score were calculated with the following equation:  $Z = \frac{Y - \hat{Y}}{RMSE}$ , where  $Y$  indicated the original cognitive score in 2011, 2013, and 2015, respectively,  $\hat{Y}$  indicated the predicted population mean cognitive score,  $RMSE$  represented the root mean square error of the regression equation [13]. In this study, the adjusted  $Z$  score were used for trajectories analysis.

The Growth Mixed Model (GMM) was used to explore the heterogeneity of cognitive trajectories, which could divide populations into several groups based on the differences in growth trajectories. Previous studies suggested that a latent growth curve model (LGCM) and latent class growth model (LCGM) should be used to explore the shape of growth curve and the number of potential trajectory classes before GMM analysis [14–15]. When the optimal LCGM model was selected, the GMM model was fitted subsequently. For model

selection, statistical indices and interpretability are often considered. Statistical indices include sample-size adjusted Bayesian information criteria (SABIC), entropy, Vuong-Lo-Mendell-Rubin likelihood ratio test (VLMR-LRT), bootstrapped likelihood ratio test (BLRT), proportion of the smallest class, average posterior probability (APP). SABIC is an information criterion with a more reduction, representing an improvement of model. Entropy is a measure of classification accuracy, ranging from 0 to 1. The larger the entropy, the better the trajectories classification. VLMR-LRT and BLRT compare the results of the  $k-1$  class model with  $k$  class model. A significant  $p$  value indicates that  $k$  class model is better than  $k-1$  class model. Besides, each trajectory class must contain enough samples, no less than 5% of total population. For APP, it is recommended to be greater than 0.7.

## Derivation and evaluation of cognitive trajectories prediction models

Logistic regression (LR) were used as the benchmark model for comparisons with another two machine learning methods, namely support vector machine (SVM) and random forest (RF). LR, a probabilistic nonlinear regression in statistical methods, were used to study the relationship between outcome and a set of factors. SVM is one of the most common ML algorithms based on maximum margin hyperplane as decision boundary. It is especially suitable for linear classification with small samples for its outputs only depend on support vectors. Meanwhile, it also performed well for nonlinear problems with kernel tricks [16]. RF, a popular ensemble learning method, is a combination of multiple decision tree classifiers. It integrates the prediction results of all decision tree classifiers to determine its final output [17]. Besides, RF is often used to assess the importance of predictors [18].

A 10-fold cross validation was used to evaluate the performance of prediction models [19]. We used accuracy to evaluate the proportion of samples classified correctly. Specifically, balanced accuracy was used in imbalanced data. The F1 score, which combines both precision and recall, was also calculated [20]. Area under the receiver operating characteristic curve (AUROC) was used to evaluate the discrimination of prediction models. Calibration was evaluated by brier score. For the above indices, 95% confidence interval (95% CI) were considered. Besides, the variable importance of predictors were plotted based on RF.

## Statistical analysis

Continuous variables were presented as mean  $\pm$  standard deviation. Categorical variables were presented as percentages. The comparisons of baseline characteristics among different trajectories were performed by appropriately choosing ANOVA test and chi square test. All the above analysis were conducted with SPSS 25.0. Trajectory class analyses were performed with Mplus 8.3 (Muthén and Muthén, 2019). Feature selection, model derivation, and model evaluation were performed with scikit-learn package in Python 3.7.6. A two-sided  $p$ -value of  $< 0.05$  was considered statistically significant.

## Results

### Heterogeneous trajectories of cognition

A linear mean trajectory was found with a negative slope for the cognition of middle-aged and elderly adults (intercept = 3.37,  $p < 0.001$ ; slope = - 0.07,  $p < 0.001$ ). The results of 2–5 classes LCGM model were presented in Table 1. A notable reduction of SABIC was observed in 2-class and 3-class. Although the results of VLMR-LRT and BLRT showed that more classes were much better ( $p < 0.05$ ), the smallest class and average posterior probability of 3-class were  $> 5.0\%$  and  $> 0.70\%$ , respectively, and its entropy was 0.64, indicating that the 3-class were more favorable. Then, GMM was used to estimate the final trajectories, and the GMM with 3-class was adopted as the final model (SABIC = 39381.36, entropy = 0.74, VLMR-LRT  $p < 0.05$ , BLRT  $p < 0.001$ ).

Table 1  
Performance of latent class growth model and growth mixture model

Model	SABIC	Entropy	VLMR-LRT $p$ -value	BLRT $p$ -value	Smallest class	APP
<b>LCGM</b>						
2-class	40054.04	0.70	0.0000	0.0000	27.7%	0.86, 0.93
3-class	39583.91	0.64	0.0000	0.0000	9.0%	0.85, 0.84, 0.80
4-class	39514.87	0.69	0.0175	0.0000	2.6%	0.80, 0.78, 0.84, 0.70
5-class	39424.36	0.63	0.0006	0.0000	3.0%	0.74, 0.81, 0.78, 0.70, 0.68
<b>GMM</b>						
3-class	39381.36	0.74	0.0206	0.0000	6.6%	0.71; 0.92; 0.72
NOTE. SABIC, sample-size adjusted Bayesian information criteria; VLMR-LRT, VUONG-LO-MENDELL-RUBIN likelihood ratio test; BLRT, Bootstrapped likelihood ratio test; APP, average posterior probability; LCGM, latent class growth model; GMM, growth mixture model.						

The cognitive trajectories were plotted (Fig. 1). The first class (83.3%) had a high cognitive score in baseline and remained at a high level subsequently (intercept = 0.19,  $p < 0.001$ ; slope = 0.018,  $p > 0.05$ ), and was named “persistently-high” (PH). The second class (6.6%) had a medium cognitive score in baseline and decreased rapidly thereafter (intercept = -0.18,  $p > 0.05$ ; slope = -0.77,  $p < 0.001$ ), and was named “medium-decrease” (MD). The third class (10.1%) had a relatively lower baseline cognitive score but showed a rapid growth subsequently (intercept = -1.453,  $p < 0.001$ ; slope = 0.353,  $p < 0.001$ ), and was named “low-increase” (LI).

## Baseline characteristics of study population

Table 2  
Baseline characteristics of the whole population and people with different trajectories

Variables	Total sample (n = 4962)	Trajectory Class			P	Missing data, NO.
		Low- increase (n = 358)	Medium- decrease (n = 234)	Persistently- high (n = 4370)		
<b>Demographics</b>						
age (y)	57.50 ± 8.25	58.25 ± 8.57	58.77 ± 9.27	57.37 ± 8.16	0.008**	0
gender, no. (%)	4947(99.7%)	357(7.2%)	233(4.7%)	4357(88.1%)	0.000***	15
male	2693(54.4%)	146(40.9%)	116(49.8%)	2431(55.8%)		
female	2254(45.6%)	211(59.1%)	117(50.2%)	1926(44.2%)		
education, no. (%)	4962(100.0%)	358(7.2%)	234(4.7%)	4370(88.1%)	0.000***	0
illiteracy	469(9.5%)	69(19.3%)	56(23.9%)	344(7.9%)		
non-illiterate	4493(90.5%)	289(80.7%)	178(76.1%)	4026(92.1%)		
marital status, no. (%)	4962(100.0%)	358(7.2%)	234(4.7%)	4370(88.1%)	0.111	0
not in married	389(7.8%)	38(10.6%)	20(8.5%)	331(7.6%)		
in married	4573(92.2%)	320(89.4%)	214(91.5%)	4039(92.4%)		
<b>Lifestyles and health behaviors</b>						
sleep time (hours)	6.52 ± 1.64	6.44 ± 2.02	6.42 ± 1.82	6.53 ± 1.60	0.407	13
having social activity, no. (%)	4962(100.0%)	358(7.2%)	234(4.7%)	4370(88.1%)	0.000***	0
no	2112(42.6%)	207(57.8%)	114(48.7%)	1791(41.0%)		
yes	2850(57.4%)	151(42.2%)	120(51.3%)	2579(59.0%)		
smoke, no. (%)	4962(100.0%)	358(7.2%)	234(4.7%)	4370(88.1%)	0.015*	0
yes	2135(43.0%)	129(36.0%)	108(46.2%)	1898(43.4%)		
no	2827(57.0%)	229(64.0%)	126(53.8%)	2472(56.6%)		
drink, no. (%)	4962(100.0%)	358(7.2%)	234(4.7%)	4370(88.1%)	0.036*	0
yes	1861(37.5%)	120(33.5%)	74(31.6%)	1667(38.1%)		

NOTE. Values are presented as mean ± standard deviation, number (%), or median (interquartile range).

\* represents  $p < 0.05$ ; \*\* represents  $p < 0.01$ ; \*\*\* represents  $p < 0.001$

Variables	Total sample ( <i>n</i> = 4962)	Trajectory Class			<i>P</i>	Missing data, NO.
		Low-increase ( <i>n</i> = 358)	Medium-decrease ( <i>n</i> = 234)	Persistently-high ( <i>n</i> = 4370)		
no	3101(62.5%)	238(66.5%)	160(68.4%)	2703(61.9%)		
<b>Cognitive &amp; depression</b>						
MMSE Score	13.31 ± 2.84	7.94 ± 2.12	12.41 ± 2.70	13.79 ± 2.40	0.000***	0
CESD-10 Score	6.97 ± 5.68	9.58 ± 6.52	8.18 ± 6.14	6.70 ± 5.52	0.000***	112
<b>Clinical information</b>						
waist circumference(cm)	82.86 ± 16.54	88.12 ± 16.96	80.17 ± 8.70	82.58 ± 16.74	0.000***	40
hypertension, no. (%)	1240(25.1%)	97(27.3%)	57(24.5%)	1086(24.9%)	0.592	18
diabetes, no. (%)	333(6.8%)	17(4.8%)	18(7.8%)	298(6.9%)	0.263	33
stroke, no. (%)	90(1.8%)	5(1.4%)	2(0.9%)	83(1.9%)	0.426	12
psychiatric problems, no. (%)	43(0.9%)	6(1.7%)	2(0.9%)	35(0.8%)	0.224	16
memory-related disease, no. (%)	46(0.9%)	7(2.0%)	4(1.7%)	35(0.8%)	0.039*	12
NOTE. Values are presented as mean ± standard deviation, number (%), or median (interquartile range).						
* represents <i>p</i> < 0.05; ** represents <i>p</i> < 0.01; *** represents <i>p</i> < 0.001						

Table 2 showed the results of baseline characteristics of the whole population and people with different trajectory classes. This study identified 4370 participants with “persistently high” class, 234 participants with “medium decrease” class, and 358 participants with “low increase” class. The mean age of the whole population was 57.50 years; and 54.4% of them were male. Of the whole population, 9.5% of them were illiteracy, 92.2% of them were married, 43.0% of them were smokers, and 37.5% of them had a history of drinking within one year. Moreover, the baseline score of MMSE and CESD-10 were 13.31 and 6.97 respectively. For the comparisons of three cognitive trajectories, significant differences were found in age, gender, education, social activity engagement, smoking, drinking, MMSE score, CESD-10 score, waist circumference, and memory-related disease (all *p* < 0.05).

## Determinants of cognitive trajectory

The results of odds ratios (OR) and its 95% CI were presented in Table 3 with binary logistic regression. Compared with “persistently-high” class, individuals with low baseline MMSE scores, memory-related disorders, younger, higher education, or stroke-free were at more risk of being classified as low-increase

class. While those with low baseline MMSE score, high CESD-10 score, memory-related disorders, the habit of smoking, low waist circumference, or illiteracy were more likely to develop a trajectory of medium-decrease.

Table 3

“Binary Logistic” odds ratios (and 95% CI) for the low-increase vs persistently-high and medium-decrease vs persistently-high

Baseline Variable	Low-increase vs. Persistently-high		Medium-decrease vs. Persistently-high	
	Odds Ratio (95% CI)	P	Odds Ratio (95% CI)	P
<b>Demographics</b>				
age	0.998 (0.997–0.999)	0.000**	1.000 (1.000-1.001)	0.407
gender	1.017 (0.998–1.036)	0.075	1.013 (0.995–1.032)	0.152
education	1.066 (1.041–1.093)	0.000**	0.937 (0.914–0.960)	0.000**
marital status	0.984 (0.960–1.009)	0.200	1.009 (0.984–1.034)	0.503
<b>Lifestyles and health behaviors</b>				
sleep time	1.003 (1.002–3.594)	0.201	1.001 (0.997–1.006)	0.501
having social activity	0.999 (0.986–1.013)	0.932	0.990 (0.977–1.003)	0.125
smoke	0.991 (0.975–1.008)	0.319	1.020 (1.003–1.038)	0.023*
drink	1.011 (0.996–1.027)	0.154	0.989 (0.974–1.005)	0.174
<b>Cognitive &amp; depression</b>				
MMSE score	0.949 (0.946–0.951)	0.000**	0.992 (0.989–0.995)	0.000**
CESD-10 score	1.000 (0.999–1.001)	0.954	1.0014 (1.0002–1.0027)	0.023*
<b>Clinical information</b>				
waist circumference	1.000 (0.999–1.001)	0.159	0.9995 (0.9991–0.9999)	0.013*
hypertension	1.001 (0.992–1.022)	0.369	0.999 (0.984–1.014)	0.888
diabetes	0.982 (0.957–1.008)	0.165	1.010 (0.984–1.036)	0.457
stroke	0.947 (0.904–0.993)	0.024*	0.955 (0.911–1.001)	0.057
psychiatric problems	1.009 (0.941–1.081)	0.810	0.988 (0.920–1.061)	0.747
memory-related disease	1.096 (1.018–1.179)	0.015*	1.084 (1.007–1.16)	0.033*
NOTE. persistently-high class was used as the reference.				

, \* represents  $p < 0.05$ ; \*\* represents  $p < 0.01$

## Feature selection

Recursive feature elimination was used for feature selection in each imputation data set of MD vs. LI, PH vs. LI, and PH vs. MD class. For MD vs. LI (Supplementary Table 1), MMSE score, waist circumference, age, CESD-10 score, sleep time, education, smoke, social activity engagement, hypertension, gender, drink, marital status were selected in the imputation data sets. Among them, MMSE score, waist circumference, age, CESD-10 score, sleep time, education, and smoke were all selected in the five data sets. For PH vs. LI (Supplementary Table 2), only two predictors were selected for all the imputation data sets, namely MMSE score, and waist circumference. For PH vs. MD (Supplementary Table 3), all the variables were selected in the data sets, while predictors including age, waist circumference, MMSE score, CESD-10 score, sleep time, social activity engagement, hypertension, drink, smoke, gender, education and marital status were the same variables in each data set.

## Performance evaluation of cognitive trajectory prediction

The performance of cognitive trajectory prediction models with the whole variables and selected variables were shown in Table 4. For prediction in the full variables set, generally, the three ML models performed well in distinguishing MD vs. LI class, and PH vs. LI class. For PH vs. LI class, the bACC, F1 score and AUROC were almost close to 0.79, 0.95 and 0.97 respectively. Meanwhile, the brier scores of the three models were close to 0.04. For MD vs. LI class, the performance of ML was also excellent with accuracy much higher than that of PH vs. LI class, and other indices were slightly lower than those of PH vs. LI class. For MD vs. PH class, the ML methods resulted in high F1 score (about 0.92) and low brier score (about 0.05) with much lower bACC (maximum: 0.51) and AUROC (maximum: 0.68). Meanwhile, logistic regression was superior to all the other ML methods ( $p < 0.05$ ) for recognizing cognitive trajectory except that in MD vs. LI (LR was equivalent to RF,  $p > 0.05$ ) when LR was used as the benchmark model.

For prediction with the selected variables, we found that ML still performed well in distinguishing MD vs. LI class and PH vs. LI class. For instance, AUROCs of MD vs. LI class and PH vs. LI class were almost close to 0.94 and 0.96 respectively. Similarly, for MD vs. PH class, ML methods had high F1 score and low brier score, but its balanced accuracy and AUROC were much lower (maximum bACC: 0.51; maximum AUROC: 0.68). Similarly, we also noted that logistic regression was superior to all the other ML methods ( $p < 0.05$ ) for recognizing cognitive trajectory except that in MD vs. LI (LR was equivalent to RF,  $p > 0.05$ ) when LR was used as the benchmark model.

Table 4  
Performance of machine learning for cognitive trajectory prediction.

	Full variables			Selected variables with RFE		
	LR	SVM	RF	LR	SVM	RF
<b>Low-increase vs Medium-decrease</b>						
Accuracy	0.88 (0.87– 0.89)	0.85 (0.84– 0.87)	0.87 (0.86– 0.88)	0.88 (0.87– 0.89)	0.86 (0.85– 0.87)	0.88 (0.87– 0.89)
F1 (95%CI)	0.84 (0.83– 0.85)	0.80 (0.78– 0.82)	0.83 (0.82– 0.85)	0.84 (0.83– 0.85)	0.83 (0.81– 0.84)	0.83 (0.82– 0.85)
Brier Score (95%CI)	0.09 (0.09– 0.10)	0.11 (0.10– 0.12)	0.11 (0.10– 0.11)	0.09 (0.09– 0.10)	0.10 (0.09– 0.11)	0.11 (0.10– 0.11)
AUROC (95%CI)	0.94 (0.93– 0.94)	0.92 (0.90– 0.93)	0.92 (0.91– 0.94)	0.94 (0.93– 0.95)	0.92 (0.91– 0.93)	0.93 (0.91– 0.94)
AUROC Comparisons ( <i>P</i> )	Reference	0.049	0.381	Reference	0.020	0.282
<b>Low-increase vs Persistently-high</b>						
bACC (95%CI)	0.79 (0.77– 0.80)	0.73 (0.71– 0.74)	0.74 (0.73– 0.76)	0.74 (0.73– 0.75)	0.76 (0.75– 0.77)	0.80 (0.79– 0.82)
F1 (95%CI)	0.95 (0.94– 0.95)	0.94 (0.94– 0.94)	0.90 (0.87– 0.94)	0.94 (0.94– 0.95)	0.95 (0.95– 0.95)	0.95 (0.95– 0.95)
Brier Score (95%CI)	0.04 (0.03– 0.04)	0.04 (0.04– 0.05)	0.06 (0.04– 0.08)	0.04 (0.03– 0.04)	0.04 (0.04– 0.04)	0.04 (0.04– 0.04)
AUROC (95%CI)	0.97 (0.96– 0.97)	0.95 (0.94– 0.95)	0.92 (0.91– 0.93)	0.96 (0.96– 0.97)	0.89 (0.88– 0.90)	0.92 (0.91– 0.93)
AUROC Comparisons ( <i>P</i> )	Reference	0.000	0.000	Reference	0.000	0.000
<b>Medium-decrease vs Persistently-high</b>						
bACC (95%CI)	0.50 (0.50– 0.50)	0.50 (0.50– 0.50)	0.51 (0.50– 0.52)	0.50 (0.50– 0.50)	0.50 (0.50– 0.50)	0.51 (0.50– 0.51)

NOTE. RFE represents recursive feature selection. bACC represents balanced accuracy. For AUROC comparisons, support vector machine, and random forest were compared with logistic regression, respectively.

	Full variables			Selected variables with RFE		
	LR	SVM	RF	LR	SVM	RF
F1 (95%CI)	0.92 (0.92– 0.93)	0.92 (0.92– 0.92)	0.90 (0.87– 0.93)	0.92 (0.92– 0.93)	0.92 (0.92– 0.92)	0.90 (0.86– 0.93)
Brier Score (95%CI)	0.05 (0.05– 0.05)	0.05 (0.05– 0.05)	0.17 (0.11– 0.23)	0.50 (0.50– 0.50)	0.50 (0.50– 0.50)	0.17 (0.11– 0.23)
AUROC (95%CI)	0.68 (0.66– 0.70)	0.60 (0.59– 0.61)	0.52 (0.49– 0.56)	0.68 (0.66– 0.70)	0.59 (0.57– 0.60)	0.52 (0.48– 0.56)
AUROC Comparisons ( <i>P</i> )	Reference	0.000	0.000	Reference	0.000	0.000
NOTE. RFE represents recursive feature selection. bACC represents balanced accuracy. For AUROC comparisons, support vector machine, and random forest were compared with logistic regression, respectively.						

## Feature importance

The importance of predictors obtained from RF were plotted for all variables (Fig. 3). For MD vs. LI and PH vs. LI, the rank of predictors was quite similar, and their consistency reached 68.75%. If the ranking difference was allowed to be within one, their consistency reached 87.50%. Moreover, the top five important predictors of MD vs. LI and PH vs. LI were the same, namely MMSE score, waist circumference, age, CESD-10 score and sleep time. In addition, the distribution of variable importance was also very similar for MD vs. LI and PH vs. LI. For PH vs. MD, the rank of variable importance was slightly different from MD vs. LI and MD vs. LI, the top five variables were age, waist circumference, MMSE score, CESD-10 score and sleep time. Also, the distribution of variable importance was significantly different from MD vs. LI and PH vs. LI. After feature selection, the ranking of variable importance of RF in different data sets was almost similar to the above results (Supplementary Fig. 1).

## Discussion

In the present study, we described the cognitive trajectories of 5-year follow-up in a nationally representative sample of community-dwelling middle-aged and older people in China and explored its potential determinants. There were three distinct trajectories of cognition, namely, persistently-high, medium-decrease, and low-increase. Compared with previous studies [21–22], similarly, the persistently-high class accounted for most of the population, indicating that majority of participants kept stable cognition through their aging periods. However, we also found an uncommon trajectory of low-increase, which was featured for low cognitive score in baseline and gradually improvement subsequently.

Compared with “persistently-high” class, individuals with low baseline MMSE scores, memory-related disorders, younger, high education, or stroke-free were at more risk of being classified as low-increase class.

While those with low baseline MMSE score, high CESD-10 score, memory-related disorders, the habit of smoking, low waist circumference, or illiteracy were more likely to have a trajectory of medium-decrease [23]. Obviously, memory-related disorders can result in low MMSE score, and the baseline MMSE score determined the cognitive trajectory to a large extent. But compared to “persistently-high” class, why people who are younger, with high education, or in stroke-free probably developed to low-increase class? It seemed out of line, we gave the following possible explanations. Younger people may suffer acute disease that resulted in low cognitive function at baseline, but after rehabilitation treatment, their cognitive function will increase in a certain. In general, people with high education have abundant knowledge, in addition, they have more ways to reinforce cognitive recovery training. Stroke is associated with an increased risk of dementia [24], deteriorate patients’ cognitive function, unluckily, the cognitive damage due to dementia is hard to reversible, so people in stroke-free have more potential in cognitive resilience. According to the above findings, we put forward some proposals to improve cognition, delay cognitive decline and promote healthy aging, such as setting up universities for the elderly to promote education, encouraging the elderly to strengthen exercise and social interaction [25]. Besides, balanced diet and nutrition, absence of smoking, good quality of sleep, reasonable weight control also mean a lot for preventing cognitive decline.

We found that ML could well distinguish cognitive trajectories, especially for MD vs. LI class and PH vs. LI class. While for MD vs. PH, ML performed less well. It implied that the existing predictors might not be suitable for prediction. We also found that the traditional logistic regression was almost superior to the ML methods in the present study, and the same results were also found in previous suicide studies [26].

The importance of the variables varied among different subgroups, but the five factors from feature selection were important for the prediction of the cognitive trajectory, which were the age, initial MMSE score [27], CESD-10 score, waist circumference, and sleep time, suggesting that these factors might be important for the prevention of cognitive decline. Previous studies have identified that cognitive function deteriorates naturally with age [28]. Depressive symptoms were associated with cognitive decline [29], so CESD-10 score can be viewed as a predictor. High waist circumference suggested abdominal obesity, nowadays, urban residents often replace walking with transportation and thereby have insufficient exercise, furthermore, the habit of drinking alcohol and unhealthy diet, and so on, which are not only associated with abdominal obesity but also have an impact on cognition [30–31]. Previous study has showed that sleep quality was negatively correlated with cognitive impairment [32–33].

This study has the following advantages: Firstly, the population was selected from a large community-based cohort study with good representation, and the predictive data was easy to obtain, the cost is low, and there is almost no risk to the population. Secondly, the mixed growth model not only takes the population heterogeneity into account, but also the differences among individuals in the population, which is helpful to dynamically understand the changing trend of cognition in middle-aged and elderly people. Finally, machine learning methods were used to predict the future cognitive trend of the population, which was helpful for the early identification of high-risk groups as well as timely personalized intervention.

## Limitations

However, the shortcomings of present study should be noted. Firstly, the sample size was relatively small considering the availability of data. Secondly, we used three waves of data (2011, 2013 and 2015) for cognitive trajectory analysis. Considering the relative short periods, we might not be able to recognize the fine cognitive changes as well as the long-term trends of cognitive development. Thirdly, machine learning methods for predicting the cognitive trajectory were only assessed in internal validation, and external validation is needed in a wider population in our future studies. Fourth, we assumed that the predictors remained unchangeable, but in fact, some variables might change with time. In future, more advanced machine learning methods, such as long short-term memory (LSTM) model with the ability for analyzing time series data [34], could be used for cognitive trajectory prediction.

## Conclusion

This study showed that three kinds of cognitive trajectories (“persistently-high”, “medium-decrease”, “low-increase”) were existed among middle-aged and elderly Chinese people. Based on easily available and low cost epidemiological data, especially for MMSE score and age, machine learning could effectively distinguish cognitive trajectories.

## Abbreviations

MMSE: Mini-mental State Examination; CESD-10: Center for Epidemiologic Studies Depression Scale-10; ML: machine learning; LR: logistic regression; SVM: support vector machine; RF: random forest; AUROC: area under the receiver operating characteristic curve; RFE: recursive feature elimination; GMM: growth mixed model; LCGM: latent class growth model.

## Declarations

### Ethical Approval and Consent to participate

Data used in study was approved by the biomedical ethics committee of Peking University, and all participants provided written informed consent.

### Consent for publication

Not applicable.

### Availability of data and materials

Data used in this study was accessible with the following link <http://charls.pku.edu.cn/>.

### Competing interests

None.

### Funding

This study was supported by the National Natural Science Foundation of China (No. 81973144).

## Authors' contributions

Y.W., M.J., C.X., S.L., Z.J. and Y.F. worked together on this article. Specifically, Y.W. conceived and designed the study. Y.W., Z.J. and C.X. contributed to the data analysis. Y.W., M.J., and C.X. drafted the manuscript. Y.F. and S.L. supervised and revised the article. All authors have approved the final article.

## Acknowledgements

We thank the China Social Science Research Center of Peking University for providing us with the data.

## Authors information

<sup>1</sup>The State Key Laboratory of Molecular Vaccine and Molecular Diagnostics, School of Public Health, Xiamen University, China. <sup>2</sup>Key Laboratory of Health Technology Assessment of Fujian Province, School of Public Health, Xiamen University, China. <sup>3</sup>National Institute for Data Science in Health and Medicine, Xiamen University, China.

## References

1. TA Salthouse. Trajectories of normal cognitive aging. *PSYCHOL AGING*. 2019;34:17-24.
2. LG Martin, RF Schoeni. Trends in disability and related chronic conditions among the forty-and-over population: 1997-2010. *DISABIL HEALTH J*. 2014;7:S4-14.
3. CA Lanea, J Hardy, JM Schott. Alzheimer's disease. *EUR J NEUROL*. 2018;25:59-70.
4. M Ansart, S Epelbaum, G Bassignana, A Bone, S Bottani, T Cattai, et al. Predicting the progression of mild cognitive impairment using machine learning: A systematic, quantitative and critical review. *MED IMAGE ANAL*. 2021;67:101848.
5. MR Arbabshirani, S Plis, J Sui, VD Calhoun. Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. *NEUROIMAGE*. 2017;145:137-65.
6. L Mosconi, M Walters, J Sterling, C Quinn, P McHugh, RE Andrews, et al. Lifestyle and vascular risk effects on MRI-based biomarkers of Alzheimer's disease: a cross-sectional study of middle-aged adults from the broader New York City area. *BMJ OPEN*. 2018;8:e19362.
7. EY Tang, O Amiesimaka, SL Harrison, E Green, C Price, L Robinson, et al. Longitudinal Effect of Stroke on Cognition: A Systematic Review. *J AM HEART ASSOC*. 2018;7.
8. L Nevin. Advancing the beneficial use of machine learning in health care and medicine: Toward a community understanding. *PLOS MED*. 2018;15:e1002708.
9. Y Zhao, Y Hu, JP Smith, J Strauss, G Yang. Cohort profile: the China Health and Retirement Longitudinal Study (CHARLS). *INT J EPIDEMIOL*. 2014;43:61-8.
10. HG Cheng, S Chen, O McBride, MR Phillips. Prospective relationship of depressive symptoms, drinking, and tobacco smoking among middle-aged and elderly community-dwelling adults: Results from the China Health and Retirement Longitudinal Study (CHARLS). *J Affect Disord*. 2016;195:136-43.

11. Q Meng, H Wang, J Strauss, KM Langa, X Chen, M Wang, et al. Validation of neuropsychological tests for the China Health and Retirement Longitudinal Study Harmonized Cognitive Assessment Protocol. *INT PSYCHOGERIATR.* 2019;31:1709-19.
12. X Wu, X Yuan, W Wang, K Liu, Y Qin, X Sun, et al. Value of a Machine Learning Approach for Predicting Clinical Outcomes in Young Patients With Hypertension. *HYPERTENSION.* 2020;75:1271-8.
13. H Li, C Li, A Wang, Y Qi, W Feng, C Hou, et al. Associations between social and intellectual activities with cognitive trajectories in Chinese middle-aged and older adults: a nationally representative cohort study. *ALZHEIMERS RES THER.* 2020;12:115.
14. P Qiu, M Zeng, W Kuang, SS Meng, Y Cai, H Wang, et al. Heterogeneity in the dynamic change of cognitive function among older Chinese people: A growth mixture model. *Int J Geriatr Psychiatry.* 2020;35:1123-33.
15. S Liu, MJ Rovine, PC Molenaar. Selecting a linear mixed model for longitudinal data: repeated measures analysis of variance, covariance pattern model, and growth curve approaches. *PSYCHOL METHODS.* 2012;17:15-30.
16. M Hearst, S Dumais, E Osman, J Platt, B Scholkopf. Support vector machines. *IEEE Intelligent Systems & Their Applications.* 1998;13:18-28.
17. Breiman. Random Forests. *MACH LEARN.* 2001;45:5-32.
18. GJ Lee, C Do, JA Suhr. Effects of personal dementia exposure on subjective memory concerns and dementia worry. *Neuropsychol Dev Cogn B Aging Neuropsychol Cogn.* 2020:1-16.
19. G Aldehim, W Wang. Determining appropriate approaches for using data in feature selection. *INT J MACH LEARN CYB.* 2017.
20. D Chicco, G Jurman. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC GENOMICS.* 2020;21:6.
21. KM Hayden, BR Reed, JJ Manly, D Tommet, RH Pietrzak, GJ Chelune, et al. Cognitive decline in the elderly: an analysis of population heterogeneity. *AGE AGEING.* 2011;40:684-9.
22. L Tu, X Lv, C Yuan, M Zhang, Z Fan, X Xu, et al. Trajectories of cognitive function and their determinants in older people: 12 years of follow-up in the Chinese Longitudinal Healthy Longevity Survey. *INT PSYCHOGERIATR.* 2020;32:765-75.
23. JW Min. A longitudinal study of cognitive trajectories and its factors for Koreans aged 60 and over: A latent growth mixture model. *Int J Geriatr Psychiatry.* 2018;33:755-62.
24. E Kuzma, I Lourida, SF Moore, DA Levine, OC Ukoumunne, DJ Llewellyn. Stroke and dementia risk: A systematic review and meta-analysis. *ALZHEIMERS DEMENT.* 2018;14:1416-26.
25. D Forbes, SC Forbes, CM Blake, EJ Thiessen, S Forbes. Exercise programs for people with dementia. *Cochrane Database Syst Rev.* 2015:D6489.
26. RC Kessler, I Hwang, CA Hoffmire, JF McCarthy, MV Petukhova, AJ Rosellini, et al. Developing a practical suicide risk prediction model for targeting high-risk patients in the Veterans health Administration. *Int J Methods Psychiatr Res.* 2017;26.

27. MD Mijajlovic, A Pavlovic, M Brainin, WD Heiss, TJ Quinn, HB Ihle-Hansen, et al. Post-stroke dementia - a comprehensive review. BMC MED. 2017;15:11.
28. N Legdeur, M Heymans, H Comijs, M Huisman, A Maier, P Visser. Age dependency of risk factors for cognitive decline. BMC GERIATR. 2018;18:187-96.
29. GS Brewster, L Peterson, R Roker, ML Ellis, JD Edwards. Depressive Symptoms, Cognition, and Everyday Function Among Community-Residing Older Adults. J AGING HEALTH. 2017;29:367-88.
30. Q Hou, Y Guan, W Yu, X Liu, L Wu, M Xiao, et al. Associations between obesity and cognitive impairment in the Chinese elderly: an observational study. CLIN INTERV AGING. 2019;14:367-73.
31. Z Liu, H Yang, S Chen, J Cai, Z Huang. The association between body mass index, waist circumference, waist-hip ratio and cognitive disorder in older adults. J Public Health (Oxf). 2019;41:305-12.
32. K Yaffe, CM Falvey, T Hoang. Connections between sleep and cognition in older adults. LANCET NEUROL. 2014;13:1017-28.
33. SK Essien, CX Feng, W Sun, M Farag, L Li, Y Gao. Sleep duration and sleep disturbances in association with falls among the middle-aged and older adults in China: a population-based nationwide study. BMC GERIATR. 2018;18:196.
34. X Hong, R Lin, C Yang, N Zeng, C Cai, J Gou, et al. Predicting Alzheimer's Disease Using LSTM. IEEE ACCESS. 2019;7:80893-901.

## Figures

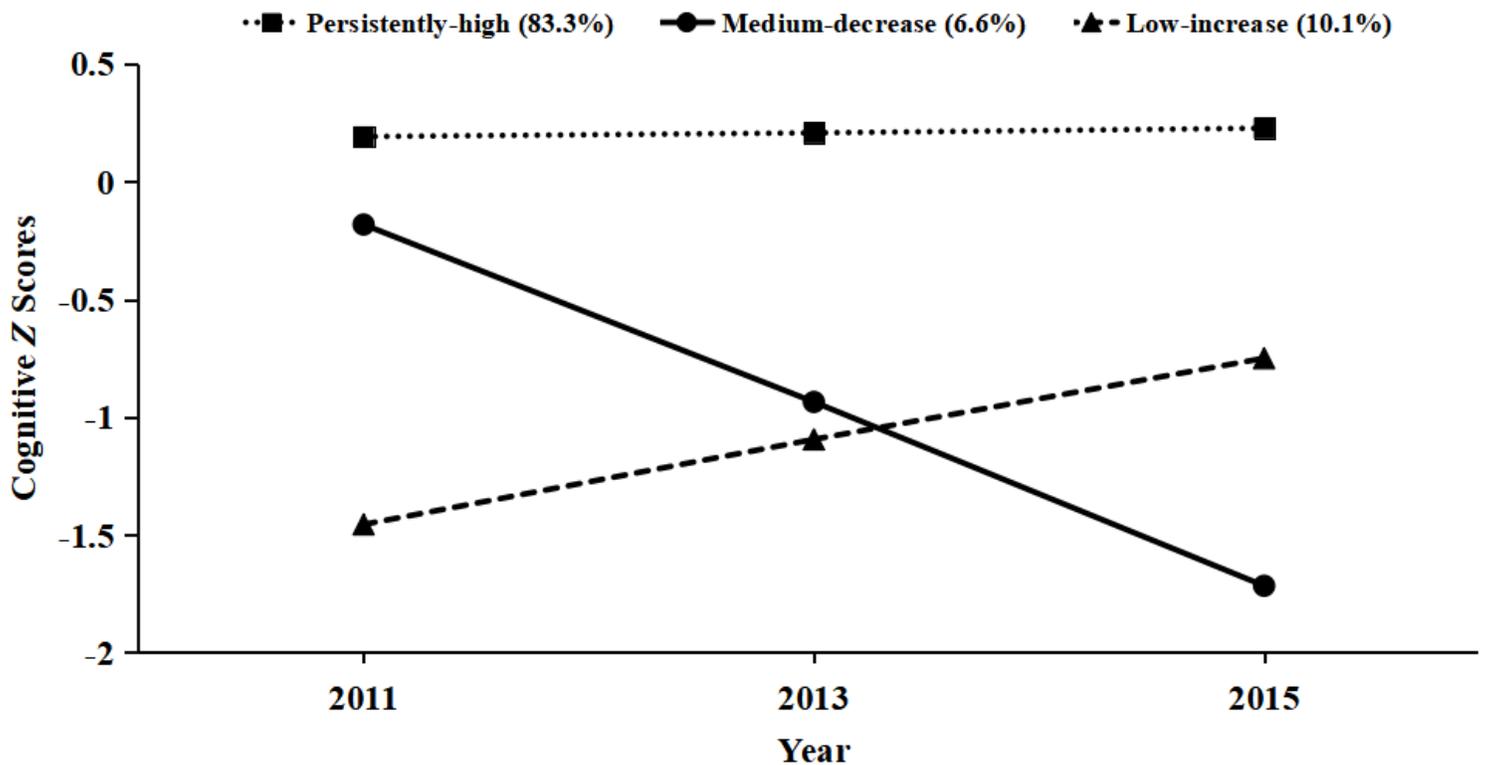
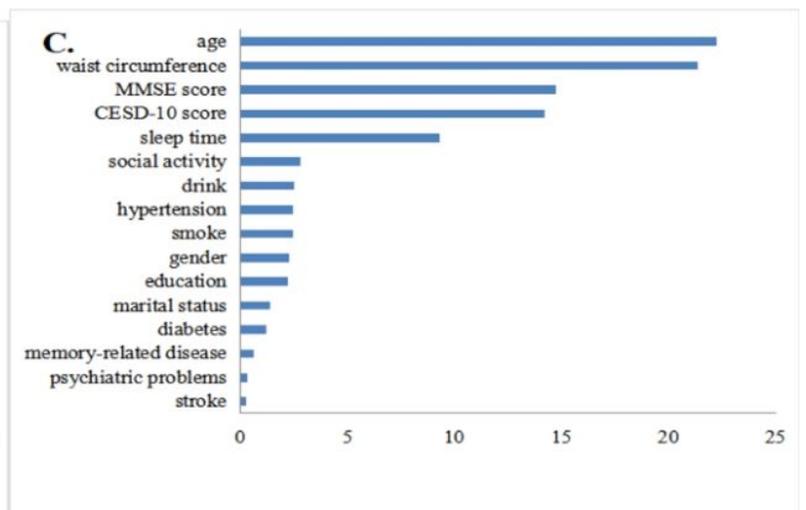
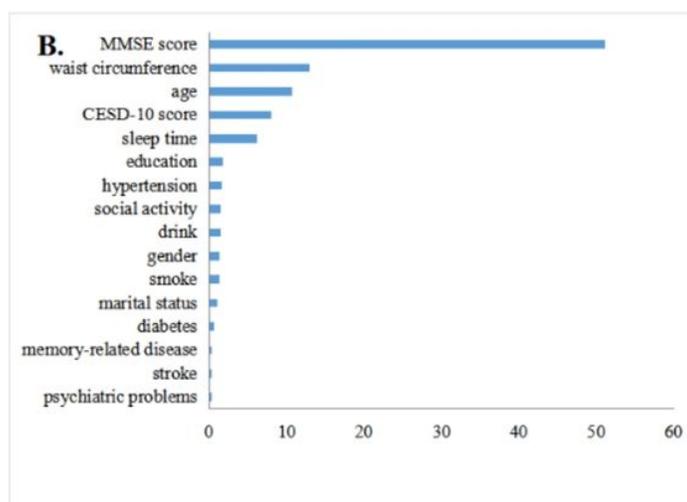
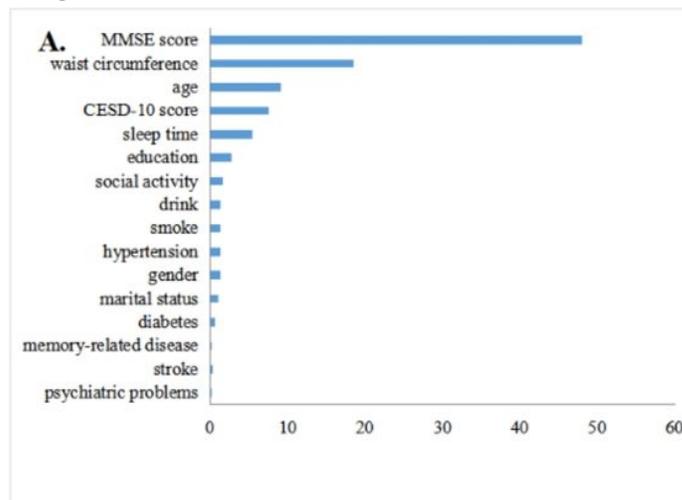


Figure 1

## Heterogeneous trajectories of cognition



**Figure 2**

Importance of all variables. Values in x-axis represents the importance of each predictors (%). A for low-increase vs. medium-decrease; B for low-increase vs. persistently-high; C for medium-decrease vs. persistently-high.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementarymaterials.doc](#)