BMC Psychiatry

RESEARCH

Open Access

Using machine learning to predict the probability of incident 2-year depression in older adults with chronic diseases: a retrospective cohort study

Ying Zheng[1], Taotao Zhang[1], Shu Yang[2], Fuzhi Wang[2], Li Zhang[1*] and Yuwen Liu[2*]

Abstract

Background Older adults with chronic diseases are at higher risk of depressive symptoms than those without. For the onset of depressive symptoms, the prediction ability of changes in common risk factors over a 2-year follow-up period is unclear in the Chinese older population. This study aimed to build risk prediction models (RPMs) to estimate the probability of incident 2-year depression using data from the China Health and Retirement Longitudinal Study (CHARLS).

Methods Four ML algorithms (logistic regression [LR], AdaBoost, random forest [RF] and k-nearest neighbor [kNN]) were applied to develop RPMs using the 2011–2015 cohort data. These developed models were then validated with 2018–2020 survey data. We evaluated the model performance using discrimination and calibration metrics, including an area under the receiver operating characteristic curve (AUROC) and the precision-recall curve (AUPRC), accuracy, sensitivity and calibrations plot. Finally, we explored the key factors of depressive symptoms by the selected best predictive models.

Results This study finally included 7,121 participants to build models to predict depressive symptoms, finding a 21.5% prevalence of depression. Combining the Synthetic Minority Oversampling Technique (SMOTE) with the logistic regression model (LR-SM) exhibited superior precision to predict depression than other models, with an AUROC and AUPRC of 0.612 and 0.468, respectively, an accuracy of 0.619 and a sensitivity of 0.546. In additiona, external validation of the LR-SM model using data from the 2018–2020 data also demonstrated good predictive ability with an AUROC of 0.623 (95% CI: 0.555– 0.673). Sex, self-rated health status, occupation, eyesight, memory and life satisfaction were identified as impactful predictors of depression.

Conclusions Our developed models exhibited high accuracy, good discrimination and calibration profiles in predicting two-year risk of depression among older adults with chronic diseases. This model can be used to identify Chinese older population at high risk of depression and intervene in a timely manner.

Keywords Chronic diseases, Risk factors, Machine learning, Depressive symptoms

Introduction

The world is experiencing a rapid acceleration of population ageing due to declining fertility and increasing life expectancy. In the coming decades, China's population ageing rate is estimated to accelerate from 11.47% to 24.71%, and Japan's ageing rate is expected to grow from 28.00% to 36.38% [1]. Particularly, this substantial demographic transition presents considerable challenges

*Correspondence:
Li Zhang
0700011@bbmu.edu.cn
Yuwen Liu
liuyw@bbmu.edu.cn
Full list of author information is available at the end of the article

for the management and treatment of chronic diseases in primary care. Approximately 75.8% of the Chinese older population suffered from at least one chronic disease, and the burden of chronic diseases has been steadily increasing over the past few years [2]. Furthermore, it is well known that there is a bidirectional association between depressive symptoms and chronic diseases [3–5]. Several studies have revealed that individuals with chronic diseases are significantly more likely to experience depression compared to those without ($P<0.001$) [6–8]. It is likely that the adverse health risk behaviors and psychobiological changes associated with depression increase the risk for chronic medical disorders. In contrast, biological changes and complications related to chronic medical disorders may precipitate depressive episodes [9].

Depression has become a growing public health concern both because of the relatively high lifetime prevalence (10–15%) and because it is associated with substantial disability [6, 10]. The number of incident cases of depression worldwide increased from 172 million in 1990 to 258 million in 2017, representing an increase of 49.86% [11]. Depression frequently co-occurs with multiple chronic diseases in complex, costly, and dangerous patterns of multimorbidity [12, 13]. In addition, depression in older adults with chronic diseases is often associated with adverse outcomes, such as cognitive impairment, causes suffering, family disruption, and disability, worsens the outcomes of many medical illnesses, and increases mortality [14]. As part of primary care for patients with chronic diseases, identifying those at high risk of depression is essential for facilitating discussions between the nursing team and the patient's family to guide interventions.

In this context, several prognostic tools using machine learning (ML) techniques have been developed to predict depression in the elderly population. Murri et al. [15] employed artificial neural network and logistic regression methods to predict the likelihood of onset of depressive symptoms in older European adults after 24-month. Many experimental studies were based on a cross-sectional design with samples consisting of both depressed and non-depressed participants without considering prior depression status information [16–18]. A 2-year community follow-up study indicated that factors influencing depression onset and persistence may differ[19]. In addition, this study design may overestimate the relevance of depressive symptoms as predictors of risk since depressive symptoms may be indices of vulnerability to depression [15, 20]. Particularly, there are few prospective studies in China based on extensive sample analyses and several ML approaches to build risk predictive models (RPMs) for predicting depression in the older population with multimorbidity. Most research on this relevant topic were systematic reviews and meta-analyses [6, 8, 12]. Therefore, RPMs based on a longitudinal design and considering the prior depression status information may more accurately identify subjects at the onset of depression and thus more effective targeting of interventions.

Building on our previous study [21] and others [15, 17, 19], the present study used an extensive existing database of older adults with chronic diseases, baseline depression informationand multiple ML methods to develop RPMs to predict depressive symptoms after 2 and identify salient factors contributing to depressive symptoms. These models can estimate the risk of developing depression over a peroid of 2 years among adults with chronic diseases who are not depressed at the time of risk assessment. Additionally, we identified critical predictors of depression by the developed best-performing models.

Methods

Data source and study population

This study followed the transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) guidelines to construct and report prediction models [22]. We utilized data from the China Health and Retirement Longitudinal Study (CHARLS), a nationally representative longitudinal survey, that contains information on social, economic and health status of Chinese people 45 years of age or older and their spouses [23]. The first national baseline survey of CHARLS was conducted in 2011 (Wave 1), and then four follow-up surveys were conducted every 2–3 years (CHARLS 2013, Wave2; CHARLS 2015, Wave3; CHARLS 2018, Wave 4; CHARLS 2020, Wave 5). All CHARLS data can be publicly available on its official website (http://charls.pku.edu.cn/). In addition, all CHARLS surveys were approved by the biomedical ethics committee of Peking University, Beijing, China (approval number: IRB 00001052–11015), and all respondents provided written informed consent.

To test broader generalizability, the CHARLS dataset was divided into two groups for different predictive steps (1) The data from the 2011–2013 and 2013–2015 surveys were used to build the depression prediction model in this study, and (2) We then used the 2018–2020 cohort data for temporal validation. For each cohort data, we applied the same inclusion and exclusion criteria to choose the study sample (Fig. S1). Participants were excluded if they met any of the following criteria: (1) Individuals with depression identified in the baseline survey; (2) Participants were under 60 years of age or without chronic disease in the baseline survey; (3) Patients with incomplete or invalid data on the 10-item Center for Epidemiological Studies Depression (CESD-10) scale

at the baseline survey and 2-year follow-up. The diagnosis of chronic diseases in the CHARLS questionnaire was based on respondents' self-reports of whether they had been diagnosed by a physician with any of the following 14 commom chronic conditions: hypertension, dyslipidemia, diabetes, cancer, chronic lung diseases, liver disease, heart disease, stroke, kidney disease, digestive diseases, psychiatric disorders, memory-related diseases, arthritis or asthma. Notably, the identification of chronic diseases was validated through repeated self-reports, but not further confirmed by clinical records or other means in the CHARLS study. Finally, 11,342 eligible individuals were included for further analysis, comprising 3,911 from the 2011–2013 cohort survey, 3,210 from the 2013–2015 cohort survey and 4,221 from the 2018–2020 cohort survey (Fig. S1).

Power analysis

To determine the required sample size for constructing RPMs, we conducted a power analysis using G*Power software (version 3.1.9.7; http://www.gpower.hhu.de/) [24]. The result indicated that a sample size of 3,495 participants was required, based on 0.9 power with an alpha of 0.05, a small effect size (odds ratio 1.2) [25] and a two-tailed test. Consequently, we had a sufficient sample to build prediction models.

Study outcome

The outcome of this study was the presence of depressive symptoms. Depressive symptoms were assessed using the CESD-10 scale in the CHARLS dataset. The CESD-10 scale has good psychometric properties to identify depressive symptoms among older Chinese adults [26–28]. The CESD-10 scale consists of 10 items that assess the frequency of certain feelings or behaviors experienced by the respondent in the past week. Each item on the CESD-10 has four response options: "rarely or none of the time (<1 day)", "some or a little of the time (1–2 days)", "occasionally or a moderate amount of the time (3–4 days)" and "most or all of the time (5–7 days)". For the eight negative items, the answers were coded as 0, 1, 2 or 3 scores, while for the two positive items, the answers were reverse-coded [29]. Thus, the total score of CESD-10 scale ranges from 0 to 30 points, with higher scores indicating more severe depressive symptoms. Consistent with previous studies [26, 30], a cutoff point of 10 was used to identify individuals with depressive symptoms. For simplicity, we used the term "depression" to refer various levels of depressive conditions with a total CESD-10 score of 10 or higher.

Candidate variables

The CHARLS survey questionnaire partially changes each survey, but some core informations are retained for long-term follow-up investigations and studies. Thus, the predictor variables and outcome identified in this study were measured in the same way in all waves of the survey to ensure procedural standardization. Based on relevant literature and available information from the CHARLS dataset, the candidate variables in this study included patient sociodemographic information, physical health history, lifestyle status, psychological status and disease treatment information. We curated a total of 49 predictors, which were systematically categorized into the following five groups: (1) For sociodemographic variables, sex, age, marital status, number of alive children and education level were included. (2) The physical health variables consist of self-rated health status, physical pain, basic (BADL) and instrumental (IADL) activities of daily living, etc. The BADL was measured using the Katz index, which consists of six activities items: bathing, dressing, toileting, transferring, continence and feeding [31]. The IADL was measured by six instrumental activities of daily living: housekeeping, preparing food, shopping, using the telephone, handling medications, and handling finances [32]. The answer for each ADL and IADL items had four options to choose: "No, I do not have any difficulty", "I have difficulty but can still do it", "I have difficulty and need help" and "I cannot do it". BADL or IADL disability is defined as responding having any difficulty with any items [32, 33]. In addition, participants who were assisted in these above activities of daily living by family members, nursing homes, paid helpers, or others were defined as having care support (otherwise no care support), as one of the predictors. (3) Lifestyle variables included alcohol consumption, smoking, sleep duration, etc. (4) Psychological variables included cognitive function and life satisfaction. The cognitive assessment tools used in CHARLS included attention, orientation, episodic memory, and visuospatial abilities [34]. Attention and orientation were assessed in the Telephone Interview for Cognitive Status (TICS) for serial subtraction of 7 from 100 (up to five times), date (year, month, day), day of the week, and season of the year; with test scores ranging from 1 to 10 [35]. Word-recall test evaluated episodic memory, which was the average number of correct immediate and delayed word recalls from a list of 10 random words [36]. The episodic memory score ranged from 0 to 10. Visuospatial ability was assessed with a figure-drawing task where participants were respondents were shown a picture of two overlapped pentagons and asked

to draw a similar figure. For the task, respondents who successfully completed the task received a score of 1 (otherwise 0) [36]. The total score of cognitive function was calculated as the sum of scores of attention, orientation, episodic memory, and visuospatial ability (range: 0–21 points). Higher scores meant better cognitive function [37]. Life satisfaction was collected using the question: "Please think about your life as a whole, how satisfied are you with it?". The responses were classified as "satisfied" or "dissatisfied." (5) Disease treatment variables were evaluated by asking individuals whether they had ever received specific treatment for their chronic disease; these treatments included medicines, chemotherapy, surgery, radiation therapy, physical therapy, occupational therapy and other relevant interventions. All predictors could be automatically matched across different wave surveys.The full list of the predictors and its encoding in prediction models are summarized in Table S1.

Data preprocessing

Due to the lack of reliable estimates, many readers will be skeptical of estimated variables with many missing values and this problem may recur in the application of the model [38]. Therefore, we excluded 14 variables with missingness greater than 30% [39, 40]. Missing values in the remaining variables were then imputed with missForest algorithm, a nonparametric random forest-based multiple imputation method for mixed-type data (missing in numeric & categorical variables) [38]. Feature selection is an important step in ML tasks, as it accelerates algorithm and improves classification accuracy by discarding irrelevant feature subsets from large and noisy raw data [41]. Elastic net, logistic least absolute shrinkage and selection operator (LASSO) regression and Boruta techniques were employed to reduce predictors on the training set. To ensure reliable results, any feature deemed irrelevant for predicting the target variable by at least two different feature selection methods was excluded. The elastic net, a regularization and variable selection method, simultaneously does automatic variable selection and continuous shrinkage to select groups of correlated variables [42]. Because the nature of the LASSO constraint it tends to produce some coefficients that are exactly 0 and hence select features that largely effect the target variable from a large and potentially multicollinear set of variables [43]. The Boruta algorithm, a wrapper approach built around a random forest classifier, which provides unbiased and stable to find all relevant features by comparing the relevance of the real features to that of the random probes [38]. The results of feature selection are presented in Tables S3 and S4. Specifically, the ten-fold cross-validation (CV) and Grid search methods were combined to find optimal hyperparameter combinations. In the logistic LASSO regression and elastic net model, features with an absolute value of the weight parameter less than 0.01 are considered as redundant or irrelevant features [44]. After filtering the 11 redundant features, 24 features were retained to construct RPMs in this study.

The class imbalance problem arose during the data processing step, which can significantly deteriorate the classification accuracy, particular with patterns belonging to the less represented classes [45]. Several techniques have been developed to solve the class imbalance issue, including Tomek Links, Synthetic Minority Oversampling Technique (SMOTE) and SMOTE + Tomek Links (SMOTETomek). To select the best model, we combined ML algorithms with three resampling techniques to develop RPMs in this study.

Model development and comparison

The machine learning workflow for this study is illustrated in Fig. 1. Four supervised ML algorithms were applied to build model: logistic regression (LR), AdaBoost, random forest (RF) and k-nearest neighbor (kNN). The 2011–2015 cohort data was randomly split into training (70%) and testing (30%) set. A ten-fold CV and grid search strategy were conducted to identify optimal hyperparameters for each model in the training set, with the area under the receiver operating characteristic (ROC) curve (AUROC) as the evaluation metrics. The final hyperparameter set of ML models are presented in Table S5. We evaluated the model performance using discrimination and calibration metrics, which are the most common metrics to assess the performance of binary classification problems. Discrimination measures the model's ability model to distinguish between different classes, including accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), the AUROC and the area under a precision-recall (PR) curve (AUPRC). Calibration, the extent to which the predicted risk scores accurately estimate the observed values, was visually assessed by a calibration plot [18]. To examine the generalization of our developed models, we performed temporal validation using a separate cohort (2018–2020 cohort data). Understanding the reasons behind a model's predictions is as important as the accuracy of the prediction in many real-world applications. Therefore, we directly calculated the variable importance for each optimal base model.

Statistical analysis

Parametric continuous variables are presented as means (standard deviations) and compared using the Student *t*

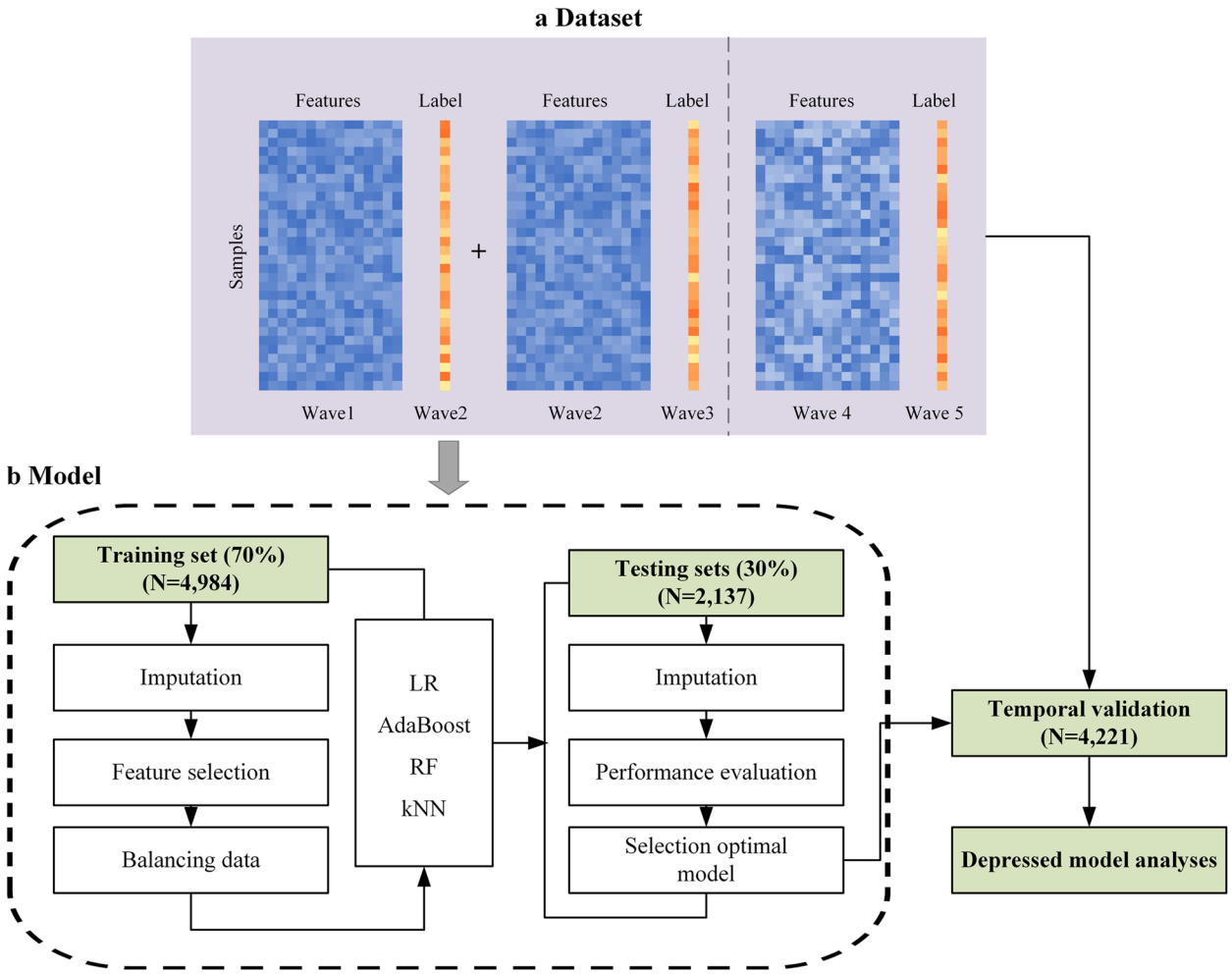

Fig. 1 Machine learning workflow in this study

test; non-parametric continuous variables are reported as median (interquartile range [IQR]) and compared using Wilcoxon rank-sum test [46]. Categorical variables are presented as frequencies (percentages) and compared using the Chi-square test. The 95% confidence intervals (CIs) were calculated using 1,000 bootstrap resamples with replacement. The logistic LASSO regression was performed using the R package Glmnet (version 4.1–8) [47]. The remaining analyses were conducted in Python (version 3.8) [48].

Results

Sample characteristics

A total of 7,121 eligible individuals were selected to develop the RPMs in this study. Besides, data from the CHARLS survey shows that the overall prevalence of special chronic diseases among the elderly has declined over the years, particularly hypertension, stomach disease and arthritis. However, an increasing number of older adults with chronic diseases are experiencing depression (Fig. S2). Table S2 describes sample characteristics categorized according to depression status (2011–2015 cohort data). Most patients received pension (84.1%), did smoke (76.4%), were married (78.6%), engaged in agricultural work (75.9%), had without experiencing physical pain (88.6%) and approximately half were male (53.2%). 1,533 (21.5%) patients have experienced depression after 2 years follow-up. Among them, the mean age was 73.5 years and 882 (57.5%) were female.

Female participants are more likely to suffer from depression compared to male participants (57.5% vs. 42.5%, $P<0.001$). Additionally, Participants with poor lifestyle and physical status had a higher proportion of depression than those with healthier lifestyle and physical status. There were significant differences in sex, education level, occupation, physical pain, hearing, memory and cognitive function between depressed and non-depressed patients ($P<0.001$, Table S2). Besides,

individuals without depression symptoms at baseline were more likely to develop depression during the 2-year follow-up (Fig. S3).

Model performance and variable importance

Table 1, Fig. 2 and Fig. 3 illustrate the performance of all ML models. Overall, models trained on imbalanced data exhibited better performance in predicting majority class (no-depression group), as indicated by higher specificity and NPV. However, after balancing the data, most classifiers demonstrated a significant improvement in accuracy for predicting minority class (depression group) compared to models trained on imbalanced data.

Compared to other algorithms, the LR-SM model exhibited better accuracy in predicting both positive and negative class. The LR-SM model had marginally good discrimination and calibration profiles (Fig. 2 and Fig. 3). In addition, the LR-SM model achieved high AUROC and AUPRC values in the validation set, which means it has good generalizability (Table 1). Therefore, we selected the LR-SM model as the optimal model to predict depression in this study (accuracy: 0.619; sensitivity: 0.546; specificity: 0.639; PPV: 0.594; AUROC: 0.612; AUPRC: 0.468).

We chose the best-performing model of each base ML model to calculate variable importance scores. Table 2 shows the top 15 important features in predicting depression (ranked from most to least important). The results of the feature importance analysis showed that sex, self-rated health status, occupation, eyesight, physical pain, and marital status were significantly associated with depression in Chinese elderly people with chronic diseases.

Table 1 Performance of machine learning models in the testing and validation set

Model	Testing set ($n=2,137$)							Validation set ($n=4,211$)	
	Accuracy	Sensitivity	Specificity	PPV	NPV	AUROC (95% CI)	AUPRC (95% CI)	AUROC (95% CI)	AUPRC (95% CI)
LR	0.784	0.033	0.990	0.484	0.789	0.632 (0.503, 0.820)	0.336 (0.262, 0.461)	0.659 (0.562, 0.681)	0.444 (0.419, 0.468)
LR-SM	0.619	0.546	0.639	0.594	0.837	0.612 (0.565, 0.618)	0.468 (0.438, 0.499)	0.623 (0.555, 0.673)	0.574 (0.564, 0.584)
LR-TL	0.784	0.033	0.990	0.469	0.789	0.629 (0.504, 0.519)	0.355 (0.258, 0.458)	0.664 (0.526, 0.640)	0.410 (0.377, 0.444)
LR-ST	0.609	0.521	0.757	0.137	0.845	0.613 (0.566, 0.626)	0.463 (0.453, 0.668)	0.578 (0.548, 0.667)	0.211 (0.171, 0.376)
AdaBoost	0.781	0.013	0.995	0.375	0.786	0.632 (0.499, 0.509)	0.300 (0.177, 0.431)	0.667 (0.520, 0.731)	0.410 (0.374, 0.444)
AdaBoost-SM	0.783	0.015	0.994	0.412	0.725	0.631 (0.499, 0.511)	0.319 (0.191, 0.451)	0.673 (0.519, 0.630)	0.456 (0.418, 0.494)
AdaBoost-TL	0.781	0.026	0.987	0.364	0.787	0.616 (0.499, 0.514)	0.300 (0.215, 0.390)	0.632 (0.499, 0.609)	0.300 (0.177, 0.431)
AdaBoost-ST	0.655	0.470	0.706	0.305	0.829	0.589 (0.563, 0.613)	0.444 (0.411, 0.477)	0.550 (0.511, 0.534)	0.359 (0.342, 0.376)
RF	0.786	0.013	0.998	0.600	0.787	0.613 (0.500, 0.512)	0.413 (0.235, 0.567)	0.626 (0.501, 0.505)	0.351 (0.264, 0.437)
RF-SM	0.647	0.502	0.686	0.305	0.834	0.617 (0.569, 0.622)	0.457 (0.423, 0.488)	0.599 (0.472, 0.687)	0.223 (0.205, 0.242)
RF-TL	0.786	0.022	0.995	0.556	0.788	0.626 (0.502, 0.515)	0.394 (0.261, 0.512)	0.679 (0.503, 0.509)	0.396 (0.326, 0.462)
RF-ST	0.629	0.528	0.657	0.297	0.835	0.611 (0.566, 0.616)	0.463 (0.431, 0.496)	0.578 (0.453, 0.668)	0.211 (0.196, 0.226)
kNN	0.786	0.012	0.995	0.991	0.786	0.615 (0.500, 0.628)	0.610 (0.601, 0.620)	0.643 (0.500, 0.703)	0.363 (0.220, 0.506)
kNN-SM	0.640	0.448	0.285	0.692	0.820	0.614 (0.543, 0.694)	0.426 (0.393, 0.459)	0.633 (0.541, 0.562)	0.374 (0.350, 0.396)
kNN-TL	0.784	0.007	0.999	0.750	0.786	0.617 (0.500, 0.627)	0.485 (0.111, 0.617)	0.637 (0.500, 0.704)	0.353 (0.251, 0.470)
kNN-ST	0.648	0.324	0.736	0.252	0.799	0.630 (0.504, 0.654)	0.361 (0.328, 0.393)	0.535 (0.525, 0.547)	0.359 (0.341, 0.377)

Note: *AUROC* the area under the receiver operating characteristic curve, *CI* Confidence interval, *PPV* Positive predictive value, *NPV* Negative predictive value, *AUPRC* The area under the precision-recall curve, *LR* Logistic regression, *RF* Random forest, *kNN* k-nearest neighbor, *SM* Synthetic Minority Oversampling Technique, *TL* Tomek Links, *ST* SMOTETomek

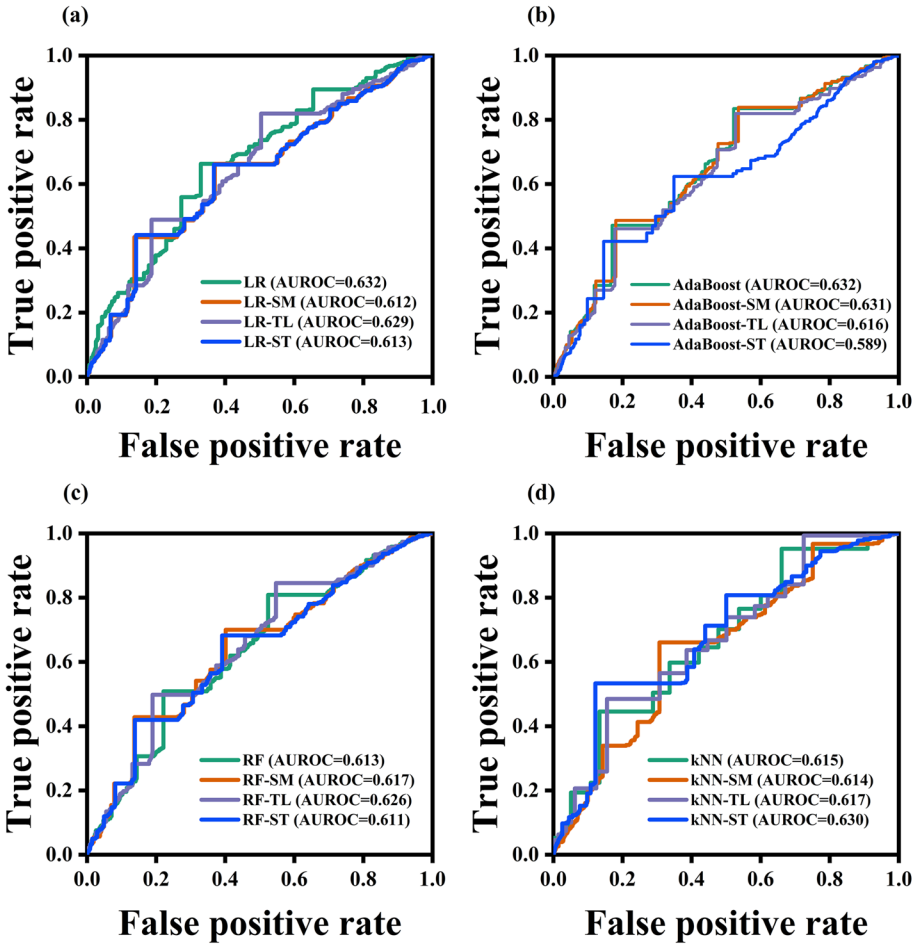

Fig. 2 The receiver operating characteristic curves for each model

Discussion

We used data from CHARLS to build models in predicting depression, including 11,342 eligible patients for the final analysis. We trained these models using a combination of four ML algorithms and three data balancing techniques. The observed depression prevalence was 21.5% among the older adults with chronic diseases. The present study examined relationships between depression and predictors by univariable analysis. Sex, education level, occupation, physical pain, hearing, and cognitive function were significantly associated with depressive symptoms in Chinese chronic diseases patients ($P < 0.001$). In addition, the LR-SM model achieved satisfactory accuracy in predicting depression. Our findings demonstrated that health-related factors was the most significant predictor for depression among older adults with chronic diseases.

This study showed that the overall prevalence of depressive symptoms among older adults with chronic diseases was 21.5%, which was different from previously reported in other countries [6, 49, 50]. The reason for this discrepancy may be the use of different diagnostic tools and settings. A systematic review and meta-analysis revealed that the prevalence of depression according to the Patient Health Questionnaire (PHQ-9) was 38.8%, 9.3–23.0% according to the International Classification of Diseases (ICD) codes, and 34.2% and 14.8% at Hospital Anxiety and Depression Scale (HADS) thresholds of 8 and 11, respectively [6, 50]. In the CHARLS dataset, the diagnosis of depression was based on self-reported questionnaires, so it cannot be ruled out that recall bias may either overestimate or underestimate the true prevalence. In a systematic analysis by Anderson et al. [5] of 42 studies reviewed, found that the prevalence of depression was assessed by self-report questionnaires (31%) than by standardized diagnostic interviews (11%). This could be because the CESD-10 scale defines a symptomatic rather than a diagnostic condition, which might have identified more sub-threshold or mild cases compared to some clinically diagnosed tools [17].

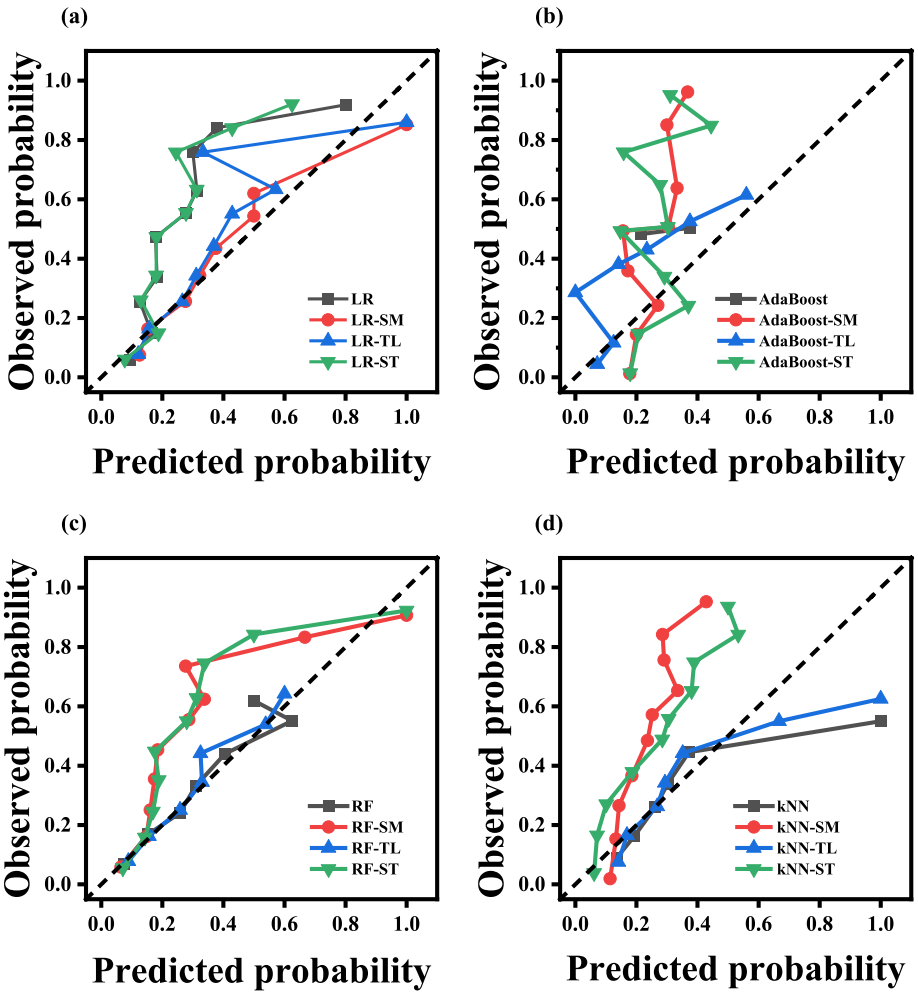

Fig. 3 Calibration for all models

Some of the top 15 most important features in this study have been previously identified in other epidemiological studies. For example, self-rated health status (SRH), the second most important feature of the LR-SM model, has been suggested as a strong predictor for depression among the elderly in many studies under several conditions [7, 51, 52]. In addition, Huang et al. [7] found that compared with the elderly with good self-rated health, those with poor self-rated health had higher risk for depression (RR: 2.40, 95% CI: 1.94–2.97). However, some studies conducted the conclusion that health status was not associated with depression in the elderly [53–55]. A systematic review and meta-analysis showed that the odds ratio (OR) of poor SRH as a function of increased depression was non-significant (OR: 1.8, 95% CI: 0.5–12.8) [53]. It is not clear whether the SRH status should be considered as a concomitant phenomena of depression or independent risk factors for increased depression.

Limitations

Our study has several limitations. First, our study was based on the longitudinal design. Thus, there was inevitably attrition due to mortality, loss of follow-up and invalid data information. In addition, this may lead to potential issue of longitudinal data bias, leading to different conclusions [56]. Second, this study focused on the prediction of depression over 2 years among older people with chronic diseases. Therefore, this study does need to fully consider the complex interrelationships change over different follow-up times and how respondents adjust to these changes in the ageing process. These important findings should be further validated with a long-term or short-term follow-up. Third, our study viewed depression as a binary classification, yet it did not further subdivide depression into more detailed categories to build models. Fourth, our study design does not consider controlling or accounting for possible confounding variables that could influence the development of depression, such as socioeconomic status,

Table 2 Top 15 Features in descending order of importance using each optimal base model for predicting depression

LR-SM	AdaBoost-ST	RF-TL	kNN-SM
1. Memory	1. Sex	1. Sex	1. Sleep duration
2. Self-rated health status	2. Marital status	2. Sleep duration	2. Physical pain
3. Public housing fund	3. Sleep duration	3. Quantity of treatment diseases	3. Received inpatient care
4. Life satisfaction	4. Household registration	4. Co-morbid conditions	4. Eyesight
5. Occupation	5. Social activity	5. Life satisfaction	5. Household registration
6. Eyesight	6. Received inpatient care	6. Household registration	6. IADL disability
7. IADL disability	7. Eyesight	7. Physical pain	7. Social activity
8. Experience of falling	8. Alcohol consumption	8. Memory	8. Occupation
9. Smoking	9. Co-morbid conditions	9. Occupation	9. Major misfortune injury experience
10. Alcohol consumption	10. Hearing	10. Received inpatient care	10. Self-rated health status
11. Major misfortune injury experience	11. Quantity of treatment diseases	11. IADL disability	11. Arthritis treatment
12. Arthritis treatment	12. Public housing fund	12. Eyesight	12. Quantity of treatment diseases
13. Sex	13. Arthritis treatment	13. Marital status	13. Alcohol consumption
14. Physical pain	14. Memory	14. Self-rated health status	14. Co-morbid conditions
15. Marital status	15. Smoking	15. Social activity	15. Sex

Abbreviations: *IADL* Instrumental activities of daily living, *LR* Logistic regression, *RF* Random forest, *kNN* k-nearest neighbor, *SM* Synthetic Minority Oversampling Technique, *TL* Tomek Links, *ST* SMOTETomek

access to healthcare or family support systems. Finally, many predictors were measured through self-reported, which may affect the accuracy and trandferability of our developed models.

Conclusions

In summary, we applied several key steps to build RPMs, including feature selection, data balancing, model performance and temporal validation. Finally, ML models based on data from the CHARLS achieved satisfactory accuracy in predicting depression among the Chinese elderly population. The discriminative ability of these ML models required validation in other national populations to examine model generalizability. In addition, our finding emphasizes that the development of depression was influenced by a myriad of variables, including sex, socioeconomic status, family support systems and physical health. Healthcare practitioners and policymakers for depression and prevention will require incorporation of this complexity situation.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12888-024-06299-6.

Supplementary Material 1

Acknowledgements
This study is based on the China Health and Retirement Longitudinal Study (CHARLS). The authors would like to thank the CHARLS research team, the field team and everyone who contributed to this study.

Authors' contributions
YZ, LZ and YL conceived and designed the study. YZ, TZ and SY analyzed and interpreted data. YZ, TZ, SY and FW drafted the manuscript. YZ, TZ and SY contributed substantial revisions to the manuscript draft. All authors read and approved the final manuscript.

Funding
This work was funded by the Talent Cultivation Program of Bengbu Medical University, grant number by51202201 and 51201209, and the Incubation Project of National Social Science Fund of Bengbu Medical University, grant number 2023byfy141sk.

Data availability
The datasets analyzed during the current study are available in the CHARLS repository (http://charls.pku.edu.cn/).

Declarations

Ethics approval and consent to participate
This study was carried out based on data extracted from the CHARLS public database, and all methods were performed according to the relevant guidelines and regulations. Written informed consent was obtained from all participants or their legal agents before the commencement of any study process. The CHARLS survey was approved by the biomedical ethics committee of Peking University, Beijing, China (approval number: IRB00001052-13074), and all respondents provided written informed consent.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]School of Nursing, Bengbu Medical University, Bengbu, China. [2]School of Health Management, Bengbu Medical University, Bengbu, China.

Received: 1 July 2024 Accepted: 14 November 2024
Published online: 02 December 2024

References

1. Chen R, Xu P, Song P, Wang M, He J. China has faster pace than Japan in population aging in next 25 years. BioSci Trends. 2019;13:287–91.
2. Wang L, Chen Z, Zhang M, Zhao Z, Huang Z, Zhang X, et al. Study of the prevalence and disease burden of chronic disease in the elderly in China. Zhonghua Liu Xing Bing Xue Za Zhi. 2019;40:277–83.
3. Herrera PA, Campos-Romero S, Szabo W, Martínez P, Guajardo V, Rojas G. Understanding the relationship between depression and chronic diseases such as diabetes and hypertension: A grounded theory study. Int J Environ Res Public Health. 2021;18:12130.
4. Clarke DM, Currie KC. Depression, anxiety and their relationship with chronic diseases: a review of the epidemiology, risk and treatment evidence. Med J Aust. 2009;190:S54-60.
5. Anderson RJ, Freedland KE, Clouse RE, Lustman PJ. The prevalence of comorbid depression in adults with diabetes: a meta-analysis. Diabetes Care. 2001;24:1069–78.
6. Moussavi S, Chatterji S, Verdes E, Tandon A, Patel V, Ustun B. Depression, chronic diseases, and decrements in health: results from the World Health Surveys. Lancet. 2007;370:851–8.
7. Huang C-Q, Zhang X-M, Bi-Rong, Dong Z-C, Lu, Yue J-R, Liu Q-X. Health status and risk for depression among the elderly: a meta-analysis of published literature. Age Ageing. 2010;39:23–30.
8. Read JR, Sharpe L, Modini M, Dear BF. Multimorbidity and depression: a systematic review and meta-analysis. J Affect Disord. 2017;221:36–46.
9. Katon WJ. Epidemiology and treatment of depression in patients with chronic medical illness. Dialogues Clin Neurosci. 2011;13:7–23.
10. Lépine J-P, Briley M. The increasing burden of depression. Neuropsychiatr Dis Treat. 2011;7(Suppl 1):3–7.
11. Liu Q, He H, Yang J, Feng X, Zhao F, Lyu J. Changes in the global burden of depression from 1990 to 2017: Findings from the Global Burden of Disease study. J Psychiatr Res. 2020;126:134–40.
12. Birk JL, Kronish IM, Moise N, Falzon L, Yoon S, Davidson KW. Depression and multimorbidity: Considering temporal characteristics of the associations between depression and multiple chronic diseases. Health Psychol. 2019;38:802–11.
13. Fiske A, Wetherell JL, Gatz M. Depression in older adults. Annu Rev Clin Psychol. 2009;5:363–89.
14. Alexopoulos GS. Depression in the elderly. Lancet. 2005;365:1961–70.
15. Murri MB, Cattelani L, Chesani F, Palumbo P, Triolo F, Alexopoulos GS. Risk prediction models for depression in community-dwelling older adults. Am J Geriatr Psychiatry. 2022;30:949–60.
16. Lin S, Wu Y, He L, Fang Y. Prediction of depressive symptoms onset and long-term trajectories in home-based older adults using machine learning techniques. Aging Ment Health. 2023;27:8–17.
17. Xiao W, Liu Y, Huang J, Huang L-A, Bian Y, Zou G. Analysis of factors associated with depressive symptoms in stroke patients based on a national cross-sectional study. Sci Rep. 2024;14:9268.
18. Shi M, Yang A, Lau ES, Luk AO, Ma RC, Kong AP, et al. A novel electronic health record-based, machine-learning model to predict severe hypoglycemia leading to hospitalizations in older adults with diabetes: A territory-wide cohort and modeling study. PLos Med. 2024;21:e1004369.
19. Harris T, Cook DG, Victor C, DeWilde S, Beighton C. Onset and persistence of depression in older people—results from a 2-year community follow-up study. Age Ageing. 2006;35:25–32.
20. Bogner HR, Morales KH, Reynolds CF III, Cary MS, Bruce ML. Course of depression and mortality among older primary care patients. Am J Geriatr Psychiatry. 2012;20:895–903.
21. Zheng Y, Zhang C, Liu Y. Risk prediction models of depression in older adults with chronic diseases. J Affect Disord. 2024;359:182–8.
22. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): The TRIPOD Statement. BMC Med. 2015;162. https://doi.org/10.1136/bmj.g7594
23. Zhao Y, Hu Y, Smith JP, Strauss J, Yang G. Cohort profile: the China health and retirement longitudinal study (CHARLS). Int J Epidemiol. 2014;43:61–8.
24. Kang H. Sample size determination and power analysis using the G* Power software. J Educ Eval Health Prof. 2021;18. https://doi.org/10.3352/jeehp.2021.18.17
25. Faul F, Erdfelder E, Buchner A, Lang A-G. Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. Behav Res Methods. 2009;41:1149–60.
26. Andresen EM, Malmgren JA, Carter WB, Patrick DL. Screening for depression in well older adults: Evaluation of a short form of the CES-D. Am J Prev Med. 1994;10:77–84.
27. Chen H, Mui AC. Factorial validity of the center for epidemiologic studies depression scale short form in older population in China. Int Psychogeriatr. 2014;26:49–57.
28. Ni Y, Tein J-Y, Zhang M, Yang Y, Wu G. Changes in depression among older adults in China: A latent transition analysis. J Affect Disord. 2017;209:3–9.
29. Li Y, Wu Y, Zhai L, Wang T, Sun Y, Zhang D. Longitudinal association of sleep duration with depressive symptoms among middle-aged and older Chinese. Sci Rep. 2017;7:11794.
30. Boey KW. Cross-validation of a short form of the CES-D in Chinese elderly. Int J Geriatr Psychiatr. 1999;14:608–17.
31. Wallace M, Shelkey M. Katz index of independence in activities of daily living (ADL). Urol Nurs. 2007;39:93–4.
32. Lawton MP, Brody EM. Assessment of older people: self-maintaining and instrumental activities of daily living. Gerontologist. 1969;9:179–86.
33. He C, Chen H, Guo L, Xu L, Liu Q, Zhang J, et al. Prevalence and factors associated with comorbid depressive symptoms among people with low back pain in China: a cross-sectional study. Front Psychiatry. 2022;13:922733.
34. Zhou S, Song S, Jin Y, Zheng Z-J. Prospective association between social engagement and cognitive impairment among middle-aged and older adults: evidence from the China Health and Retirement Longitudinal Study. BMJ Open. 2020;10:e040936.
35. Li M, Wang N, Dupre ME. Association between the self-reported duration and quality of sleep and cognitive function among middle-aged and older adults in China. J Affect Disord. 2022;304:20–7.
36. Li J, Cacchione PZ, Hodgson N, Riegel B, Keenan BT, Scharf MT, et al. Afternoon napping and cognition in Chinese older adults: findings from the China health and retirement longitudinal study baseline assessment. J Am Geriatr Soc. 2017;65:373–80.
37. Yao Y, Wang K, Xiang H. Association between cognitive function and ambient particulate matters in middle-aged and elderly Chinese adults: Evidence from the China Health and Retirement Longitudinal Study (CHARLS). Sci Total Environ. 2022;828:154297.
38. Kursa MB, Rudnicki WR. Feature selection with the Boruta package. J Stat Softw. 2010;36:1–13.
39. Stekhoven DJ, Bühlmann P. MissForest—non-parametric missing value imputation for mixed-type data. Bioinformatics. 2012;28:112–8.
40. Ikeda T, Cooray U, Hariyama M, Aida J, Kondo K, Murakami M, et al. An interpretable machine learning approach to predict fall risk among community-dwelling older adults: a three-year longitudinal study. J Gen Intern Med. 2022;37:2727–35.
41. Li J, Cheng K, Wang S, Morstatter F, Trevino RP, Tang J, et al. Feature selection: a data perspective. ACM Comput Surv. 2017;50:1–45.
42. Zou H, Hastie T. Regularization and variable selection via the elastic net. J R Stat Soc Ser B-Stat Methodol. 2005;67:301–20.
43. Tibshirani R. Regression shrinkage and selection via the lasso. J R Stat Soc Ser B-Stat Methodol. 1996;58:267–88.
44. Zheng X, Wang F, Zhang J, Cui X, Jiang F, Chen N, et al. Using machine learning to predict atrial fibrillation diagnosed after ischemic stroke. Int J Cardiol. 2022;347:21–7.
45. Barandela R, Valdovinos RM, Sánchez JS, Ferri FJ: The imbalanced training sample problem: Under or over sampling? In: Structural, Syntactic, and Statistical Pattern Recognition: Joint IAPR International Workshops, SSPR 2004 and SPR 2004, Lisbon, Portugal, August 18–20, 2004. Proceedings: 2004: Springer; 2004: 806–814.
46. Nanayakkara S, Fogarty S, Tremeer M, Ross K, Richards B, Bergmeir C, et al. Characterising risk of in-hospital mortality following cardiac arrest using machine learning: A retrospective international registry study. PLos Med. 2018;15:e1002709.
47. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw. 2010;33:1–22.
48. Srinath K. Python–the fastest growing programming language. Int J Eng Technol. 2017;4:354–7.

49. Kennedy GJ, Kelman HR, Thomas C. The emergence of depressive symptoms in late life: the importance of declining health and increasing disability. J Community Health. 1990;15:93–104.
50. Matcham F, Rayner L, Steer S, Hotopf M. The prevalence of depression in rheumatoid arthritis: a systematic review and meta-analysis. Rheumatology. 2013;52:2136–48.
51. Schoevers R, Beekman A, Deeg D, Geerlings M, Jonker C, Van Tilburg W. Risk factors for depression in later life; results of a prospective community based study (AMSTEL). J Affect Disord. 2000;59:127–37.
52. Unsar S, Sut N. Depression and health status in elderly hospitalized patients with chronic illness. Arch Gerontol Geriatr. 2010;50:6–10.
53. Cole MG, Dendukuri N. Risk factors for depression among elderly community subjects: a systematic review and meta-analysis. Am J Psychiat. 2003;160:1147–56.
54. Cho HJ, Lavretsky H, Olmstead R, Levin MJ, Oxman MN, Irwin MR. Sleep disturbance and depression recurrence in community-dwelling older adults: a prospective study. Am J Psychiat. 2008;165:1543–50.
55. De Beurs E, Beekman A, Geerlings S, Deeg D, Van Dyck R, Van Tilburg W. On becoming depressed or anxious in late life: similar vulnerability factors but different effects of stressful life events. Br J Psychiatry. 2001;179:426–31.
56. Mehrabi N, Morstatter F, Saxena N, Lerman K, Galstyan A. A survey on bias and fairness in machine learning. ACM Comput Surv. 2021;54:1–35.

Publisher's Note