RESEARCH

A hybrid machine learning model of depression estimation in home-based older adults: a 7-year follow-up study

Shaowu Lin[1,2,3†], Yafei Wu[1,2,3†] and Ya Fang[1,2,3*]

Abstract

Background: Our aim was to explore whether a two-step hybrid machine learning model has the potential to discover the onset of depression in home-based older adults.

Methods: Depression data (collected in the year 2011, 2013, 2015 and 2018) of home-based older Chinese ($n =$ 2,548) recruited in the China Health and Retirement Longitudinal Study were included in the current analysis. The long short-term memory network (LSTM) was applied to identify the risk factors of participants in 2015 utilizing the first 2 waves of data. Based on the identified predictors, three ML classification algorithms (i.e., gradient boosting decision tree, support vector machine and random forest) were evaluated with a 10-fold cross-validation procedure and a metric of the area under the receiver operating characteristic curve (AUROC) to estimate the depressive outcome.

Results: Time-varying predictors of the depression were successfully identified by LSTM (mean squared error $=0.8$). The mean AUCs of the three predictive models had a range from 0.703 to 0.749. Among the prediction variables, self-reported health status, cognition, sleep time, self-reported memory and ADL (activities of daily living) disorder were the top five important variables.

Conclusions: A two-step hybrid model based on "LSTM+ML" framework can be robust in predicting depression over a 5-year period with easily accessible sociodemographic and health information.

Keywords: Machine learning, LSTM, Depression, Home-based elderly, Prediction

Introduction

Depression is the most common psychological problem in older people. Almost 7% of older adults worldwide are currently suffering from depressive disorder [1]. Globally, the COVID-19 pandemic has added a 27.6% increase of depression [2]. Geriatric depression usually impairs an individual's life style and even physical functioning, and could severely affect the work and daily life of older

adults [3]. In China, the prevalence of 11.5% to 21.1% in depression gave a more sever challenge [4]. Thus, the need of estimating depression earlier and accurately is more and more urgent.

However, depression forecasting is considerably difficult and dominated by various risk factors. Several attempts have been made to predict depression utilizing regression models such as generalized linear regression, and machine learning techniques with demographic characteristics, social factors and health data. For instance, Tsai Y F et al. compared the severity of depression and its main influencing factors among nursing home elderly in different regions, in which a total of 214 older adults from Hong Kong and 150 older adults

†Shaowu Lin and Yafei Wu contributed equally to this work.

*Correspondence: fangya@xmu.edu.cn

[1] The State Key Laboratory of Molecular Vaccine and Molecular Diagnostics, School of Public Health, Xiamen University, Xiamen 361102, China
Full list of author information is available at the end of the article

Lin *et al. BMC Psychiatry*   (2022) 22:816

Page 2 of 13

from Taiwan were included. Logistic regression analysis suggested that life satisfaction, gender, income, and self-reported health had a great impact on depression prediction of nursing home elders. While for Hong Kong elders, significant predictors of depression were functional status, cognition, and life satisfaction [5]. Furthermore, Chang et al. used the multinomial logistic regression to explore the depression development trajectory of the elderly in Taiwan by gender stratification, and the results found that old aged men with less social support had a higher burden of depression. However, due to the limitations of traditional regression methods in handling high-dimensional and non-linear data, their performance is often limited [6]. Also, compared with traditional regression approaches, machine learning-based models have a superior predictive ability [7]. Several studies using ML algorithms (e.g., RF) have found significant improvements in the accuracy of predicting depression [8, 9]. Helen et al. [10] used decision tree to uncover non-linear associations and interactions among physical and mental health factors, as well as cognition and magnetic resonance imaging in depressed older adults through investigating a total of 81 participants (51 major depressive disorder patients, 30 controls). They found that executive function and cognitive tests of verbal fluency had the best influencing prediction for late-life depression. Also, the results further suggested the direct association between depression and cognition. Aris Supriyanto [11] used C4.5 algorithm to estimate the risk of postpartum depression, and they found the greatest profit on the blood pressure, psychological variables, body temperature, indicating the greater impact of the three factors on the depressed individuals, and should be prioritized for intervention and treatment. Md. Rafiqul Islam et al. [12] performed depression analysis on 7145 Facebook data obtained from an Australia online social media platform. To investigate identification of depression, they proposed three machine learning techniques, namely decision trees, k-nearest-neighbor, support vector machine. The results showed that the predictive performance (accuracy) of support vector machine was superior to other ML approaches (99%, 96%, 88%). In addition to the above commonly used methods, association analysis, frequent pattern trees, artificial neural networks, and ensemble learning (such as GBDT: gradient boosting decision tree) have also been used in depression risk research [13]. For example, Thanathamathee et al. [14] compared adolescent depression with a consideration of severity (light, medium, and severe), the study found that AdaBoost algorithm significantly improved the performance of constructing model (accuracy: 82.7%). Additionally, in order to explore the physiological characteristics of depressed patients in depth, Yusra Ghafoor et al. [9] applied frequent pattern

tree and association analysis on depression database containing 5,964 records and found that sleep disorders, sleepless nights or sleep too much, easily fatigue are the most important clinical symptom of depressed patients. Although the above classifiers were powerful in prediction tasks, their performance was still limited due to the lack of capturing temporal information in the traditional features. Thus, a modern deep learning algorithm of long short-term memory network (LSTM) was proposed to capture the time-sensitive features, which could be conveniently used to predict the future stage of the disease. For the purpose of improving the prediction of depression in the long-term period, Xu et al. developed a deep neural network framework (LSTM) for the individualized prediction of depressive disorder based on 22-year longitudinal community dwelling data among older adults in the United States (AUC=0.87) [15].

Depression is marked by a chronic and longer-term course. To date, numerous studies have been published using longitudinal cluster analysis to examine the course of depression and examine heterogeneity in trajectories of depression. Some studies also have investigated the determinants of its heterogenous course. Indeed, a temporal relation can be found among these factors. Studies on the determinants of depression have shown that such risk factors may evolve over time. Yet, previous studies on the association between depression and predictors largely dominated by features at a single wave that usually cannot take time-varying features into account, and thus they may not fully reflect how dynamic and stability of these features have an affect on depression. Thus, exploring the time-dependent relationship between these risk factors can help medical workers better predict the risk of depression. Moreover, studies estimating depression for population of home-based older adults are scarce, but home-based eldercare will remain the main selection for older Chinese in a long-term period of future, due to the ethics of Chinese "filial piety" [16].

To the best of our knowledge, few studies have focused on the prediction of depression among home-based older adults in the next few years using longitudinal data. In the current study, we aim to propose a two-step hybrid machine learning model for depression classification and investigate the following: 1) are the LSTM algorithm a suitable model to predict the level of different depression risk factors in home-based older adults in the next two years by capturing time series informations; 2) how do three ML classifiers compare to classical regression model in terms of predicting the onset of depression in the home-based elderly population in two years; 3) which features are important predictors of depression for home-based older adults in a large Chinese community cohort based on the material of demographic, social

Lin *et al. BMC Psychiatry* (2022) 22:816

Page 3 of 13

economy, lifestyle and health status, and can new features be recognized in addition to established risk factors?

Methods

Study design and participants

This study was a secondary analysis of data obtained in the China Health and Retirement Longitudinal Study (CHARLS). Specific details of this nationwide study have been reported in previous studies [17]. The protocol had received permission from the Biomedical Ethics Committee of Peking University, and an institutionally informed consent form was signed by all participants. 2548 home-based older adults (60+ years) with community dwelling data obtained in the year of 2011, 2013, 2015 and 2018 were used in the current study, in which participants with a CESD-score >10 in wave 1, 2 and 3 were excluded (Fig. 1). A summary of missing values of all predictors were listed in Supplementary Table 1. The missing values were replaced utilizing the multiple imputation method, which is based on 5 duplicates and a chained equation, using the R "mi" package. Besides, a sensitivity analysis was conducted only on participants with complete data to evaluate the robustness of our models.

Predictor variables

In our study, 22 variables were included as candidate predictors. Specifically, 24 predictors firstly derived from the previous forecasting study of depression in literature [18–20], especially machine learning (ML) research [19, 21], were selected as prioritized. Secondly, we selected the same predictors as in the wave 1-3 due to the difference in the variable structure of the questionnaires under different waves and thus the variable of major misfortune injury experience was removed. Similarly, the variable of CESD-10 score was also not included as a categorical predictor since the participants with a CESD-score >10 in wave 1, 2 were excluded. Thirdly, we reduced the number of candidate predictors to a proposed event per variable (EPV) value of 10 (i.e., 10 cases per predictor) [22] to overcome methodological limitation, such as selecting unimportant variables. Finally, a set of 22 variables (divided into three groups) in the year 2011 were selected as predictors (Supplementary Table 4). Details of these variables are presented as follows:

(a) For demographic variables, geographical location (eastern vs. central vs. western) [23], age [24], sex [20], rural/urban community [25], marital status [26] were included, and age and sex were regarded
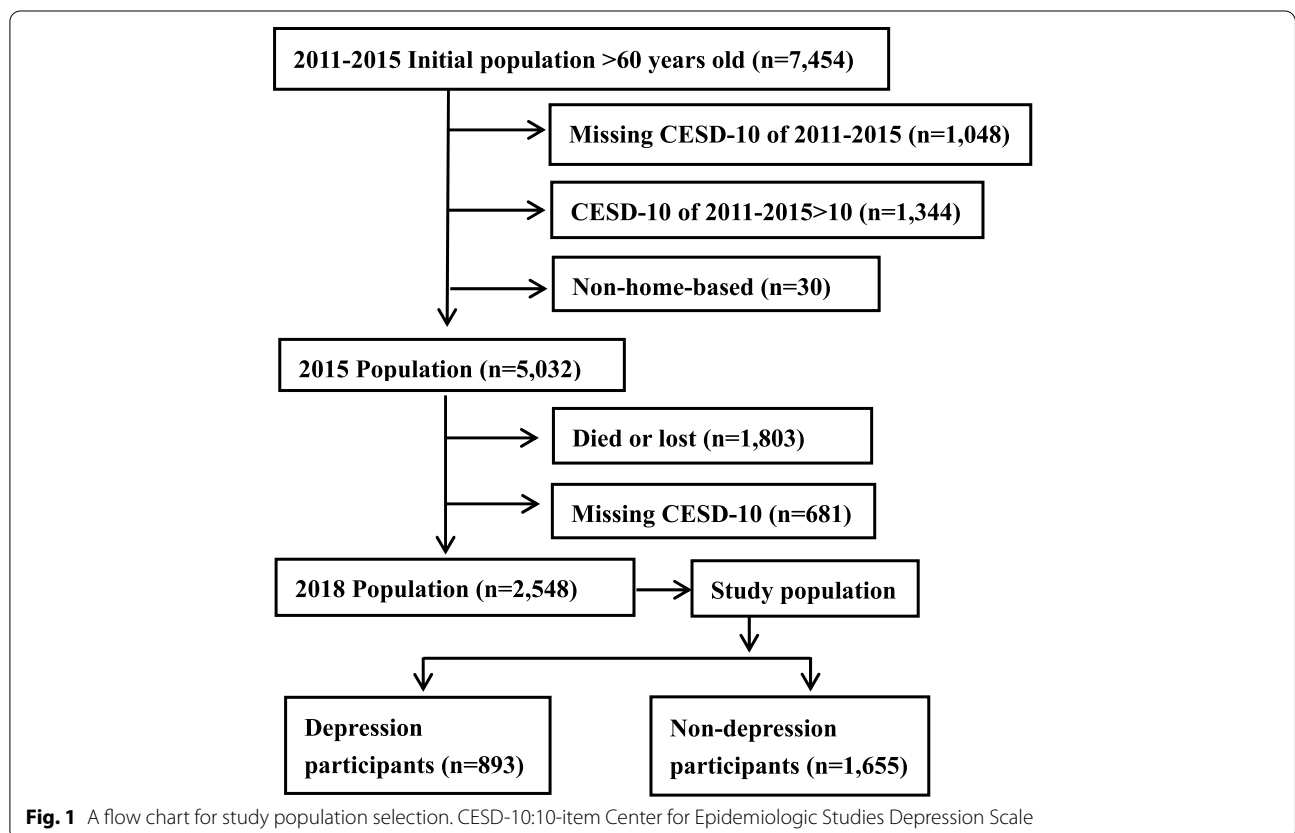

Fig. 1 A flow chart for study population selection. CESD-10:10-item Center for Epidemiologic Studies Depression Scale

Lin *et al. BMC Psychiatry*   (2022) 22:816

Page 4 of 13

as auxiliary input because they are variables that are not needed to predict over time. Geographical location was divided into eastern, central, and western regions according to the 2011 China health statistics yearbook. Marital status was categorized as married (married/partnered), and single (never married/divorced/separated and widowed).

(b) For socioeconomic variables, we considered educational level [27], household per capita income [28], household registration [29], occupational status [30], medical insurance (yes/ no) [31]. Education level was dichotomized as low-level (elementary school and below) and high-level (middle school and above). Household per capita income was defined as total household income divided by the number of people living in the family, and was grouped into three categories based on an interval of 5000 RMB. Household registration was categorized as agriculture, non-agriculture and not registered. Occupational status was divided into agricultural work, non-agricultural work, retired and unemployed/never work [23].

(c) For variables in lifestyle and health status, cognitive ability [32], sleeping time [33], self-rated memory [34], life satisfaction [35], ADL (activities of daily living) disorder [36], self-reported health status [37], social activities in the past month [38], smoking [39], alcohol drinking [27], chronic diseases [40], disability [41], medical services experience in the past month [42] were selected as predictors. Consistent with previous studies [43], cognitive ability was evaluated by episodic memory and mental intactness, and the global cognitive scores were calculated as the sum of the scores of episodic memory and mental intactness with a range from 0 to 21 [44]. In order to better observe the cognitive ability, the total cognition score was divided into two categories: above-average cognitive scores (high: score>10.5) and under-average cognitive scores (low: score<10.5) [45]. ADL impairment was measured by asking participants whether they had any difficulties in taking a bath, eating, getting in and out of bed, dressing, using the toilet, defecating, doing housework, cooking, making phone calls, taking medicine, shopping and managing finances due to health and memory problems in the past 3 months, which was previously reported by Katz [46] and Lawton [47]. If any difficulty was reported, the participant would be classified as having difficulties in ADL. Social activities experience (over the past month) covers interaction with friends, voluntary or charity work, stock investment, and other 8 kinds of social activity. Participants who had been

diagnosed with hypertension, dyslipidemia, diabetes or high blood glucose, cancer, chronic lung diseases, and other 9 kinds of chronic disease were defined as having chronic disease. Medical services experience refers to a participant who have visited a hospital, or doctor's practice, or been visited by a doctor for outpatient care in the last month. Major misfortune injury experience refers to a participant who had ever been injured in a traffic accident or any other major accidents.

Outcome variables

The 10-item Center for Epidemiologic Studies Depression Scale (CESD-10, Supplementary Table 5) was applied for assessing depression symptoms during the 4 waves of survey. The reliability and validity of CESD-10 has been validated in Chinese older population, Cronbach's alpha = 0.86 [48]. Respondents were asked about the number of days they experienced with different emotions during the past week, and a respondent who had a CESD-10 score of at least 11 was defined as suffering depression in our study [49].

Data analysis

The depression prediction task mainly consisted of two steps (Fig. 2). In the first stage, we performed the LSTM algorithm to estimate the value of risk factors of participants in 2015 utilizing the first 3 waves of data, which constituted a raw dataset. More specially, we used the categorical predictors in waves 1-2 to predict the corresponding factors of depression for elderly in wave 3 (2015) using a randomly data separation method, that is, 70% of the samples for training the LSTM model, and the remaining 30% of the samples, a validation dataset, was used for prediction to ouput the predicted factors in 2015. In the second stage, three commonly used machine learning (ML) approaches were applied to investigate whether the predicted features of 2015 could accurately predict depression in wave 4 (2018), and the output factors from the validation dataset were combined with the depression results from wave 4 to form a new dataset, in which 70% of the data was used for model construction, and the remaining 30% of data was used for model testing. Model performance was evaluated by metrics of accuracy, sensitivity, positive predictive value (PPV) and AUROC. In addition, brier score was selected as the calibration index. The construction and evaluation of machine learning methods are completed by python 3.7 with Pytorch and Scikit-learn toolkit. A bilateral *p* value of <0.05 was considered statistically significant. The feature importance was explored by calculating Shapley
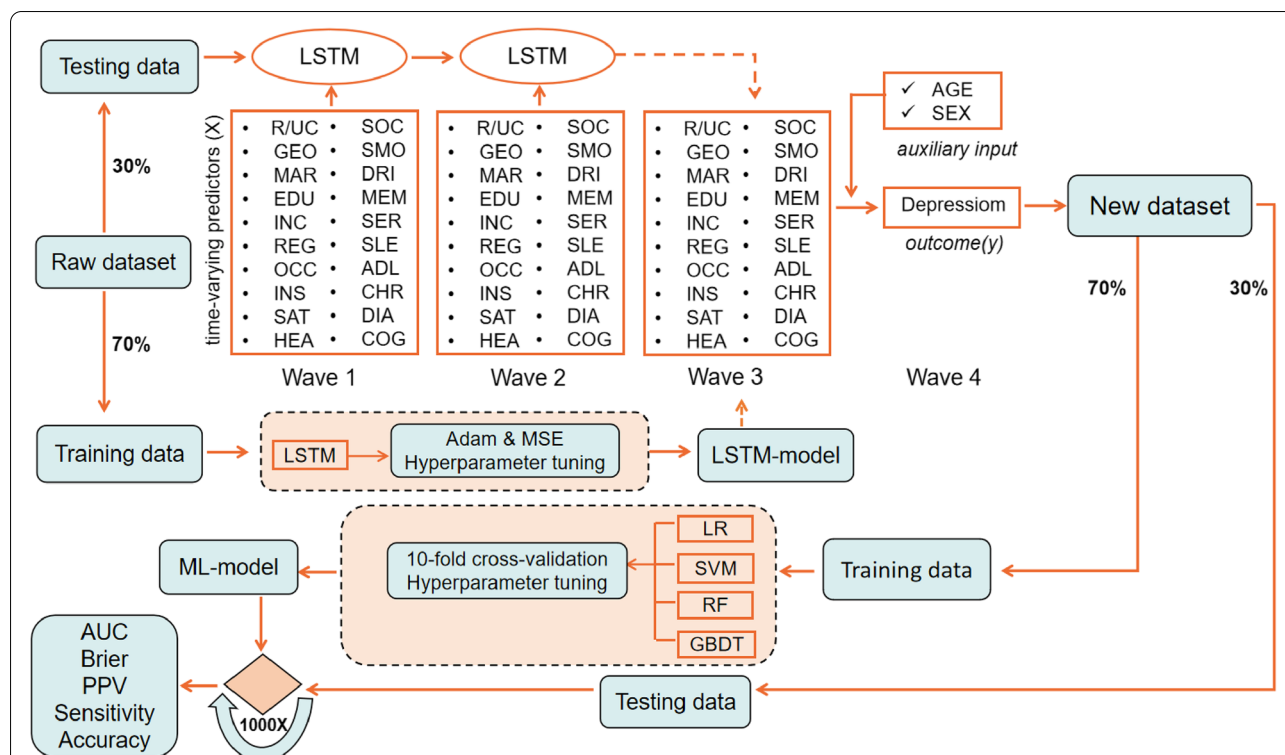
Lin *et al. BMC Psychiatry* (2022) 22:816

Page 5 of 13

Fig. 2 The architecture of the hybrid model for estimation of depression. Raw dataset used Wave 1-3 of depression data, and the output predicted data of Wave 3 and the outcome of wave 4 were constructed into a new dataset. MSE: mean squared error; ML: machine learning; RF: random forest; GBDT: gradient boosting decision tree; SVM: support vector machines; LR; logistic regression. R/UC: rural/urban community, GEO: geographical location, MAR: marital status, EDU: education level, INC: household per capita income, REG: household registration, OCC: occupation status, INS: medical insurance, SAT: life satisfaction, HEA: self-reported health status, SOC: social activities, SMO: smoking, DRI: drinking, MEM: self-rated memory, SER: medical service, SLE: sleeping time, ADL: ADL disorder, CHR: Chronic disease, DIA: Disability, COG: Cognitive ability

values via the SHAP package (v. 0.39.0) and visualized by beeswarm plot.

Five prediction algorithms were employed in our study: the binary logistic regression model (LR), a typical conventional statistical model, and three commonly ML models, including random forests (RF), gradient boosting decision tree (GBDT), and support vector machines (SVM); and a time-series based deep learning method (long short-term memory network: LSTM). SVM is a controlled classification algorithm based on statistical learning theory. The core of SVM is based on the concept of finding the most appropriate decision function that separates the two distinct classes on the basis of the definition of hyperplane, which can distinguish the two classes from each other in a most appropriate way [50, 51]. GBDT is a flexible and non parametric statistical learning technique for classification and regression. It improves the prediction appearance results by gradually improving the estimation [52]. RF is an ensemble algorithm that inputs the same test dataset to all learned decision trees and collected the results by the majority vote method. It has high prediction power for high-order

data, which have a large number of explanatory variables and complicated interactions between them [53]. Since the above models are traditional statistical and machine learning models, both of which are shallow learning models, and cannot capture the interdependence of predictors in different bands of longitudinal data during the prediction process. Therefore, we proposed LSTM algorithm, a special kind of recurrent neural network (RNN), to capture the interdependence of predictors in different waves of longitudinal data and estimate the future trend of predictors. LSTM, with the chain of repeating neural network modules, is designed to avoid the long-term dependency problem, which might be able to connect previous information to the present task and encode the temporal relation between features [54]. The adjustment of the hyperparameters is essential for model construction. For LSTM model, the learning rate is one of the most important parameters that can directly affect the convergence of the model. We repeatedly tried the value of learning rate range from 1e-1 to 1e-4 during the training process. Adam optimizer and MSE (mean squared error) of loss function were used. We evaluated the

Lin *et al. BMC Psychiatry*   (2022) 22:816

Page 6 of 13

training performance of LSTM by looking at the training curve, and a steadily decreasing loss value was observed in Supplementary Figure 1. Considering that the fitting result of the training set would affect the generalization ability of the test set, once the test error stopped falling or the error started to increase, we have to stop training. The hyperparameter settings are shown in Supplementary Table 6. A standard machine learning technique is trained on the training set, and then were evaluated on the test set. For each analysis, we randomly split the whole data into training (70% of the entire sample) and testing datasets (30%) [55]. We deployed a standard machine learning protocol with 10-fold cross-validation [56], hyperparameter tuning, using the training dataset. The detailed information about hyperparameter tuning strategy of three ML algorithms was listed in Supplementary Table 7. In the model evaluation, the accuracy, sensitivity, PPV (positive predictive value), and AUROC of the discrimination metrics were selected to evaluate the prediction performance of the proposed models. In addition, brier score was used to reflect the calibration of models and all the testing process was repeated 1000 times to take the average of those outputs as a stable estimation (Fig. 2). For the RF and GBDT classifiers, we further calculated the relative importance of the predictors according to their contribution to prediction accuracy.

Results

Table 1 summarizes the 20 predicted features and 2 auxiliary predictors of 2015 for depressed and no-depressed participants in wave 4 (year of 2018) using the validation dataset. The proportion of males for elderly in no-depressed and depressed group were 56.8% and 52.6%, respectively. There were significant differences between two groups by rural/urban Community, household per capita income, self-reported health status, drinking, disability, cognitive ability. The baseline characteristics was presented in Supplementary Table 2.

Supplementary Figure 2 presents the correlation of each predicted risk factors by LSTM model. We found their correlation coefficients of all pair-wise features were lower than the value of 0.7, and the largest relation was observed between cognitive ability and chronic disease, achieved 0.64.

The forecasting performance of LSTM model is shown in Supplementary Figure 1. The mean squared error (MSE) was quite similar between the training and validation curves, both achieved around 0.8, which suggested that the time-series based model had a good fit on the risk factors forecasting in the first task Fig. 3. A shows the ROC curves of different ML models for depression prediction. Generally, RF had the best performance, with an AUC value of 0.749, and the LR was significantly lower

than those of other three ML models. RF also had the highest value of PPV (0.650). Although the PPV of SVM model (0.462) was slightly lower, the sensitivity was the highest (0.469) compared with GBDT (0.303), RF (0.206), and LR (0.264). The accuracy of RF, LR, SVM, GBDT were 0.752, 0.713, 0.701, 0.682 respectively. For the calibration metric, the scaled brier score of RF was the lowest, while the values of the other three algorithms were relatively higher, ranging from 0.183 and 0.205 (Table 2). Supplementary Table 3 also suggested the robustness of our models when a sensitivity analysis was performed only on participants with complete data. DCA results (Fig. 3. B) showed that within the threshold ranges of 0.00-0.20 and 0.28-0.40, the net benefit of RF was the largest. The SVM achieved the best net benefit within the threshold ranges of approximately 0.20-0.28 and 0.44-0.48, and within the threshold ranges of around 0.52-0.84, the net benefit of GBDT was much higher.

Cognitive ability and self-reported health status were consistently ranked as 1st or 2nd important predictors for RF and GBDT models (Fig. 4). The two classification models ranked self-reported memory and ADL disorder as two of the top five important predictors, which was mainly composed of lifestyle and health status variables. While, life satisfaction and sleep time were just ranked one of five most important predictors for GBDT and RF, respectively. For demographic variables, geographical location and rural/urban community was two of the top ten important predictors in forecasting the depressed elderly with an exception. Regarding socioeconomic variables, two models ranked occupational status as moderately important predictors (Top 15). However, household per capita income, educational level, medical insurance were consistently the least important predictors in GBDT and RF models.

Discussion

This study retrospectively collected 4 Waves of CHARLS longitudinal data, a total of 2,548 older adults, in which demographic information, socioeconomic, lifestyle variables and health status were obtained. The results showed that LSTM model could successfully predict the risk factors in the next one year. We also explored the difference of predicted features in two outcomes and the correlation of each predicted features. The founding suggested that there was a weak collinearity between these predictors. RF achieved the best performance for the prediction of depression based on the features predicted by LSTM, and our results revealed that cognitive ability, self-reported health status, self-reported memory, and ADL disorder were top 5 predictors.

In our designed predictive models, we observed the AUCs ranged from 0.639 to 0.749, and accuracies ranged

Lin *et al. BMC Psychiatry*   (2022) 22:816

Page 7 of 13

Table 1 Predicted characteristics in 2015 and univariate analysis of association with depression in 2018 based on the validation dataset

Variables	Non-depression (n=532) % or Mean (SD)	Depression (n=232) % or Mean (SD)	Test of association
Demographic variables			
Age, year			
60-	32.36	35.26	χ^2=0.51
70-	47.40	42.51	
80-	20.24	22.23	
Sex			
Male	56.77	52.59	χ^2=1.14
Female	43.23	47.41	
Rural/Urban Community			
Rural	58.83	69.40	χ^2=7.65**
Urban	41.17	30.60	
Geographical location			
Eastern	35.90	39.66	χ^2=1.91
Central	56.58	51.29	
Western	7.52	9.05	
Marital status			
Single	22.74	25.86	χ^2=0.869
Married	77.26	74.14	
Social economy variables			
Education level			
Low	91.73	92.67	χ^2=0.196
High	8.27	7.33	
Household per capita income, yuan			
<5000	0.38	2.59	χ^2=7.62**
5000-10000	99.62	97.41	
Household registration			
Agriculture	85.15	86.21	χ^2=0.95
Non-Agriculture	14.85	13.79	
Occupation status			
Agricultural work	6.02	8.19	χ^2=1.78
Non-agricultural work	77.26	77.59	
Retired	16.73	14.22	
Medical Insurance			
No	7.33	11.21	χ^2=3.12
Yes	92.67	88.79	
Life style & health status variables			
Life satisfaction			
Satisfied	9.40	8.62	χ^2=0.121
Medium	57.89	58.62	
Not satisfied	32.71	32.76	
Self-reported health status			
Good	25.38	9.91	χ^2=92.95***
Fair	55.83	37.50	
Poor	18.80	52.59	
Social activities			
Never	54.51	51.29	χ^2=0.67
Ever	45.49	48.71	

Lin *et al. BMC Psychiatry*   (2022) 22:816

Page 8 of 13

Table 1 (continued)

Variables	Non-depression (n=532)	Depression (n=232)	Test of association
	% or Mean (SD)	% or Mean (SD)	
Smoking			
Never	73.50	71.12	χ^2=0.46
Ever	26.50	28.88	
Drinking			
Never	63.53	77.16	χ^2=13.71***
Ever	36.47	22.84	
Self-rated memory			
Good	32.71	29.31	χ^2=0.96
Fair	51.32	53.02	
Poor	15.98	17.67	
Medical service			
No	81.95	80.17	χ^2=0.34
Yes	18.05	19.83	
Sleeping time, hour			
0-	15.60	13.79	χ^2=0.699
4-	25.94	26.29	
6-	27.44	29.74	
8-	31.02	30.17	
ADL disorder			
No	73.87	73.28	χ^2=0.03
Yes	26.13	26.72	
Chronic disease			
No	51.50	55.17	χ^2=0.87
Yes	48.50	44.83	
Disability			
No	90.23	80.60	χ^2=13.50***
Yes	9.77	19.40	
Cognitive ability, score			
Low	54.89	71.12	F=17.712***
High	45.11	28.88	

χ^2 denotes the Pearson's Chi square statistic; *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

from 0.682 to 0.752. Meanwhile, the mean AUCs of the three ML models were higher than that of traditional statistical model (LR) in the estimation of elderly depression symptoms. Richard Dinga et al. [57] assessed 804 individuals from the Netherlands Study of Depression and Anxiety (NESDA), demonstrated a model with C statistics of 0.66 to distinguish patients with and without a unipolar depression diagnosis at 2-year follow-up. K.-S. Na, et al. examined baseline (2016) and follow-up (2017–2018) data of the Korea Welfare Panel Study (KoWePS) to predict the future onset of depression, achieving AUCs of 0.870. When comparing our proposed models with other similar studies predicting depression on the individual level, our results fall into the upper level of the AUC range (0.65–0.84) [58]. However, only several commonly

ML algorithms were used to build predictive model in previous studies, while our proposed hybrid ML framework was scarce in the community-based longitudinal studies.

To detect features importance, the two optimal tree-based models, RF and GBDT, were used to show the interpretability of the inner mechanisms in models' decisions. Overall, cognitive ability, self-reported health status, self-rated memory, and ADL disorder were the top 5 important predictors. The association between cognitive impairment and depression is still debatable. Two studies reporting no effect of cognitive impairment on depressive population [59, 60] while two other studies reported that cognitive impairment increased risk of developing into depression [37, 61]. ADL disorder was a robust predictor
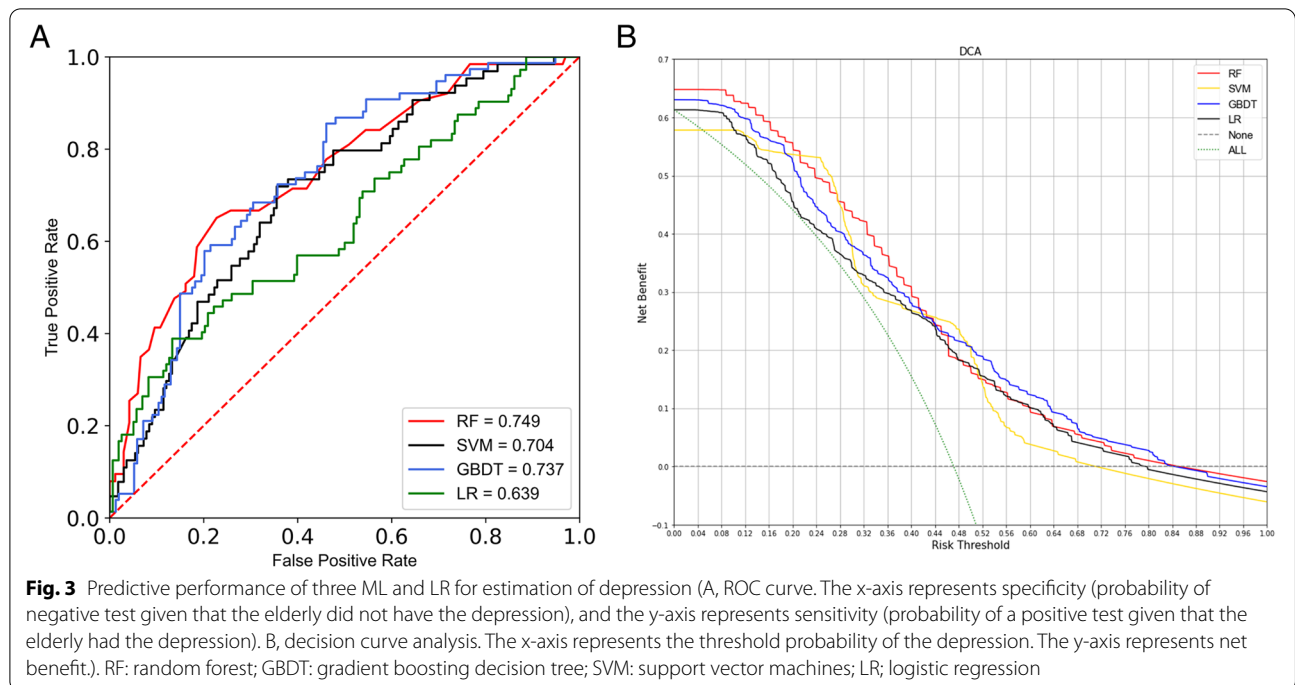
Fig. 3 Predictive performance of three ML and LR for estimation of depression (A, ROC curve. The x-axis represents specificity (probability of negative test given that the elderly did not have the depression), and the y-axis represents sensitivity (probability of a positive test given that the elderly had the depression). B, decision curve analysis. The x-axis represents the threshold probability of the depression. The y-axis represents net benefit.). RF: random forest; GBDT: gradient boosting decision tree; SVM: support vector machines; LR; logistic regression

Table 2 Model performance in predicting onset of depression

Model	Accuracy	Sensitivity	PPV	Brier score	AUC (95% CI)
SVM	0.701	0.469	0.462	0.183	0.703 (0.702-0.705)
GBDT	0.682	0.303	0.535	0.198	0.737 (0.731-0.738)
RF	0.752	0.206	0.650	0.170	0.749 (0.744-0.749)
LR	0.713	0.264	0.594	0.205	0.639 (0.633-0.640)

RF random forests, *GBDT* gradient boosting decision tree, *SVM* support vector machines, *LR* logistic regression, *PPV* Positive Predictive Value, *CI* confidence interval; Parameter optimization: RF (ntree=200, mtry=40), SVM (C=0.1, γ=0.0001), GBDT (α= 8, β=0.001, γ= 300).

of depression across a number of studies [38]. With the increase of age, older people may be confronted with more health problems that can affect their life independency. Our results provided supportive evidence that older adults having limitations in ADLs were more likely to report depression. In the lifestyle and health status variables, self-reported health status and self-rated memory were also the considerable important predictors, which is similar to the results of Kuo et al. and Liang et al. They found older adults with poor self-rated health were more likely to experience higher symptom burdens [40, 60] . All in all, the higher ranking features in the two models (e.g., cognitive ability, sleeping time, self-reported memory) could share common underlying biological mechanisms, that is, the relationship between the activation of microglia [62] and the onset of depression.

However, neither the marital status nor the chronic diseases [63] were identified as important factors in the current study. Possible explanations might be the heterogeneity of samples, we only divided the marital status into two classification and used the number of chronic diseases instead of the specific types of disease. Therefore, further studies focusing on these controversies would shed light on the mechanism of depression among the older population.

In terms of demographic variables, our main finding is that geographical location and rural/urban community were the only relatively important predictor in the prediction of depression, which is a key point for our findings. Home-based older adults who live in western region were more vulnerable to depression compared to their peers in other regions. A possible reason might be that compared with those in the eastern and middle region in China,

(See figure on next page.)

Fig. 4 Variables selected by RF and GBDT with relative relevance weights. (A/C) SHAP (SHapley Additive exPlanation) values are ordered by value of a feature to the predictions made by the RF/GBDT. The position on the x-axis on shows whether the effect of that value is associated with a higher or lower prediction for a given observation. Red color indicates the feature is high for that observation or low (blue). (B/D) Summary of mean SHAP values or overall magnitude of a feature's impact on prediction of depression by the RF/GBDT

Fig. 4 (See legend on previous page.)

Lin *et al. BMC Psychiatry*   (2022) 22:816

Page 11 of 13
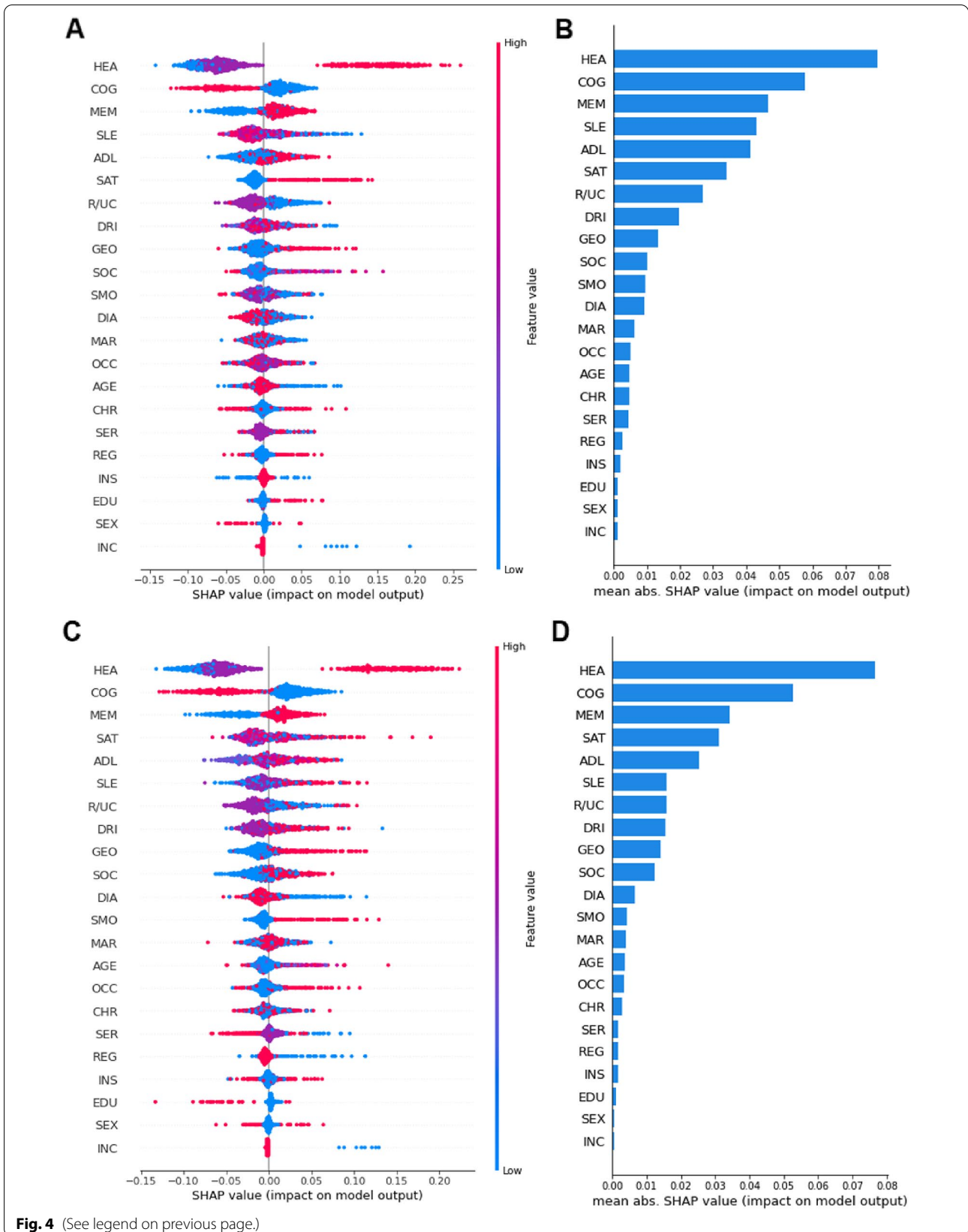
people in the west were living with less income, poorer health care support and worse living conditions [64], which would result in severer depressive condition. However, it is a pity to regard socioeconomic predictors, which showed less predictive values.

Our study could provide reference value for clinical practice. The identified risk factors can be used to inform the community prescribers, e.g., using self-reported measures along with inexpensive cognitive testing for episodic memory and mental intactness, and to target preventive interventions for improving the remission of depression. Self-reported information including sleep time, self-reported memory, CESD-10 score and geographical location. Specially, such identified risk factors and training model could be used for developing risk assessment tool (e.g., risk calculator or APP system), which likely including four modules of "individual information input", "data management", "disease prediction", and "disease intervention". Once the index data of elderly has been entered into the system, the individual development pattern of depression can transmit to community prescribers, which can help them complete the decision-making processes earlier, and adopt lifestyle intervention or clinical treatment for patients as soon as possible.

Most notably, this is the first study, to our knowledge, to investigate the hybrid ML model for depression estimation targeting a home-based population, particularly introducing a deep neural network algorithm (LSTM) to construct model based on 7-year longitudinal survey data in Chinese community-dwelling older adults. Furthermore, understanding the temporal dependence of theses predictors and what estimates a more worsening depression pattern is essential for designing customized interventions aimed to improve quality of life in these elderly people. Moreover, our hybrid model focused on the long-term course before onset of depressive, including the predictors of multiple time point, can also help the community medical providers to find people with positive symptoms as early as possible.

However, it is also important to bear in mind the limitations of the current study. Firstly, the external validation in an independent dataset was not conducted in the current analysis. Although within-sample cross-validation is known to be an almost unbiased method of population generalizability [65], it may not completely be suitable for the different characteristics of data from different samples. Validating our findings in large and independent data is the next important study. Secondly, due to limitation in data amount and follow-up waves, ML algorithms such as random forest and LSTM may not be able to represent their forecasting potential. Moreover, depression-related treatment information during the study was not collected, therefore, we did not know if the change of depression was affected by treatment. Thirdly, we converted the CESD-10 score into a binary outcome variables since our study was aimed to explore the future depression risk for those who were free of depression over the past three years, which would be better for primary screening in general population. The CESD-10 only evaluates the week before the assessment. Because of that, more detailed information to the courses and degree of the depression may be lost and even depression may not be accounted for, while for a more reliable determination of depression, more frequent evaluations and longer periods of follow-up are required. In our future research, we aim to verify our models in clinical scenarios with larger sample size, more waves of follow-up information, multi-task learning, and multimodal characteristics, to improve the accuracy of predicting model.

Conclusion

In conclusion, this study developed a two-step hybrid model based "LSTM+ML" framework in depression estimation for home-based elderly. Utilizing the easily accessible sociodemographic and health predictors, the two-step hybrid machine learning prediction showed potential to distinguish no depression from those with depression. The decision support system based on the hybrid models may be valuable for community medical providers.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12888-022-04439-4.

> **Additional file 1.**

Declarations

Author details
[1]The State Key Laboratory of Molecular Vaccine and Molecular Diagnostics, School of Public Health, Xiamen University, Xiamen 361102, China. [2]National Institute for Data Science in Health and Medicine, Xiamen University, Xiamen 361102, China. [3]Key Laboratory of Health Technology Assessment of Fujian Province, School of Public Health, Xiamen University, Xiamen 361102, China.

References
1. WHO. Depression and Other Common Mental Disorders: Global Health Estimates: Geneva: World Health Organization; 2017. p. 1–24.
2. Santomauro DF, Herrera AMM, Shadid J, Zheng P, Ashbaugh C, Pigott DM, et al. Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. Lancet. 2021;398(10312):1700–12.
3. Koh HK, Parekh AK. Toward a United States of Health: Implications of Understanding the US Burden of Disease. JAMA. 2018;319(14):1438–40.
4. Suhara Y, Xu Y, Pentland AS. Acm: DeepMood: Forecasting Depressed Mood Based on Self-Reported Histories via Recurrent Neural Networks. In:the 26th International Conference. International World Wide Web Conferences Steering Committee(WWW): 2017 Apr 3 – 7 2017; Perth, Australia; 2017. p. 715–724.
5. Tsai YF, Chung JWY, Wong TKS, Huang CM. Comparison of the prevalence and risk factors for depressive symptoms among elderly nursing home residents in Taiwan and Hong Kong. Int J Geriatr Psychiatry. 2005;20(4):315–21.
6. Chang Y-S, Hung W-C, Juang T-Y. Ieee: Depression Diagnosis based on Ontologies and Bayesian Networks. In: the 33rd IEEE International Conference on Systems, Man, and Cybernetics (SMC): 2013 Oct 13-16 2013; Manchester, England; 2013. p. 3452–3457.
7. Hatton CM, Paton LW, McMillan D, Cussens J, Gilbody S, Tiffin PA. Predicting persistent depressive symptoms in older adults: A machine learning approach to personalised mental healthcare. J Affect Disord. 2019;246:857–60.
8. Ay B, Yildirim O, Talo M, Baloglu UB, Aydin G, Puthankattil SD, et al. Automated depression detection using deep representation and sequence learning with EEG signals. J Med Syst. 2019;43(7):1.
9. Ghafoor Y, Huang Y-P, Liu S-I. An intelligent approach to discovering common symptoms among depressed patients. Soft Computing. 2015;19(4):819–27.
10. Lavretsky H, Kitchen C, Mintz J, Kim M-D, Estanol L, Kumar A. Medical burden, cerebrovascular disease, and cognitive impairment in geriatric depression: modeling the relationships with the CART analysis. CNS Spectr. 2002;7(10):716–22.
11. Supriyanto A, Suryono S, Susesno JE. Implementation Data Mining using Decision Tree Method-Algorithm C4.5 for Postpartum Depression Diagnosis. In: 3rd International Conference on Energy, Environmental and Information System (ICENIS) - Strengthening Planning and Implementation Energy, Environment, and Information System Toward Low Carbon Society: 2018 Aug 14-15 2018; Semarang, Indonesia; 2018. p. 12–15.
12. Islam MR, Kabir MA, Ahmed A, Kamal ARM, Wang H, Ulhaq A. Depression detection from social network data using machine learning techniques. Health Inform Sci Syst. 2018;6(1):8–8.
13. Cho G, Yim J, Choi Y, Ko J, Lee S-H. Review of Machine Learning Algorithms for Diagnosing Mental Illness. Psychiatry Investig. 2019;16(4):262–9.
14. Thanathamathee P, Ieee: Boosting with Feature Selection Technique for Screening and Predicting Adolescents Depression. In: 4th International Conference on Digital Information and Communication Technology and it's Applications (DICTAP): 2014 May 06-08 2014; Bangkok, Thailand; 2014: 23-27.
15. Xu Z, Zhang Q, Li W, Li M, Yip PSF. Individualized prediction of depressive disorder in the elderly: A multitask deep learning approach. Int J Med Inform. 2019;132.
16. Lin S, Wu Y, Fang Y: Comparison of Regression and Machine Learning Methods in Depression Forecasting Among Home-Based Elderly Chinese: A Community Based Study. Frontiers in Psychiatry 2022, 12.
17. 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(1):61–8.
18. C RAAB, C DCS, Mhs: What Are the Causes of Late-Life Depression? Psychiatric Clinics of North America 2013, 36(4):497-516.
19. Kessler RC, van Loo HM, Wardenaar KJ, Bossarte RM, Brenner LA, Cai T, et al. Testing a machine-learning algorithm to predict the persistence and severity of major depressive disorder from baseline self-reports. Mol Psychiatry. 2016;21(10):1366–71.
20. Richardson R, Westley T, Gariepy G, Austin N, Nandi A. Neighborhood socioeconomic conditions and depression: a systematic review and meta-analysis. Soc Psychiatry Psychiatr Epidemiol. 2015;50(11):1641–56.
21. Chekroud AM, Zotti RJ, Shehzad Z, Gueorguieva R, Johnson MK, Trivedi MH, et al. Cross-trial prediction of treatment outcome in depression: a machine learning approach. Lancet Psychiatry. 2016;3(3):243–50.
22. Studerus E, Ramyead A, Riecher-Rossler A. Prediction of transition to psychosis in patients with a clinical high risk for psychosis: a systematic review of methodology and reporting. Psychol Med. 2017;47(7):1163–78.
23. He M, Ma J, Ren Z, Zhou G, Gong P, Liu M, et al. 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;242:135–42.
24. Chan MF, Zeng W. Exploring risk factors for depression among older men residing in Macau. J Clin Nurs. 2011;20(17-18):2645–54.
25. Mansori K, Shiravand N, Shadmani FK, Moradi Y, Allahmoradi M, Ranjbaran M, et al. Association between depression with glycemic control and its complications in type 2 diabetes. Diab Metab Syndrome. 2019;13(2):1555–60.
26. Chen Y-Y, Wong GHY, Lum TY, Lou VWQ, Ho AHY, Luo H, et al. Neighborhood support network, perceived proximity to community facilities and depressive symptoms among low socioeconomic status Chinese elders. Aging Ment Health. 2016;20(4):423–31.
27. Yaroslavsky I, Pettit JW, Lewinsohn PM, Seeley JR, Roberts RE. Heterogeneous trajectories of depressive symptoms: Adolescent predictors and adult outcomes. J Affect Disord. 2013;148(2-3):391–9.
28. Librenza-Garcia D, Passos IC, Feiten JG, Lotufo PA, Goulart AC, de Souza SI, et al. Prediction of depression cases, incidence, and chronicity in a large occupational cohort using machine learning techniques: an analysis of the ELSA-Brasil study. Psychol Med. 2021;51(16):2895–903.
29. Kessler RC, Bromet EJ: The Epidemiology of Depression Across Cultures. In: Annual Review of Public Health, Vol 34. Volume 34, edn. Edited by Fielding JE; 2013: 119-138.
30. Ouyang P, Sun W. Depression and sleep duration: findings from middle-aged and elderly people in China. Public Health. 2019;166:148–54.
31. Na K-S, Cho S-E, Geem ZW, Kim Y-K. Predicting future onset of depression among community dwelling adults in the Republic of Korea using a machine learning algorithm. Neurosci Lett. 2020;721.
32. Vinkers DJ, Gussekloo J, Stek ML, Westendorp RGJ, van der Mast RC. Temporal relation between depression and cognitive impairment in old age: prospective population based study. Br Med J. 2004;329(7471):881–3.
33. Gehrman P, Seelig AD, Jacobson IG, Boyko EJ, Hooper TI, Gackstetter GD, et al. Millennium Cohort Study T: Predeployment Sleep Duration and Insomnia Symptoms as Risk Factors for New-Onset Mental Health Disorders Following Military Deployment. Sleep. 2013;36(7):1009–18.
34. Kaup AR, Byers AL, Falvey C, Simonsick EM, Satterfield S, Ayonayon HN, et al. Trajectories of Depressive Symptoms in Older Adults and Risk of Dementia. JAMA Psychiat. 2016;73(5):525–31.

Lin *et al. BMC Psychiatry*   (2022) 22:816

Page 13 of 13

35. Luoma I, Korhonen M, Salmelin RK, Helminen M, Tamminen T. Long-term trajectories of maternal depressive symptoms and their antenatal predictors. J Affect Disord. 2015;170:30–8.

36. Unsar S, Dindar I, Kurt S. Activities of daily living, quality of life, social support and depression levels of elderly individuals in Turkish society. J Pakistan Med Assoc. 2015;65(6):642–6.

37. Kuchibhatla MN, Fillenbaum GG, Hybels CF, Blazer DG. Trajectory classes of depressive symptoms in a community sample of older adults. Acta Psychiatr Scand. 2012;125(6):492–501.

38. Byers AL, Vittinghoff E, Lui L-Y, Hoang T, Blazer DG, Covinsky KE, et al. Twenty-Year Depressive Trajectories Among Older Women. Arch Gen Psychiatry. 2012;69(10):1073–9.

39. Costello DM, Swendsen J, Rose JS, Dierker LC. Risk and protective factors associated with trajectories of depressed mood from adolescence to early adulthood. J Consult Clin Psychol. 2008;76(2):173–83.

40. Liang J, Xu X, Quinones AR, Bennett JM, Ye W. Multiple Trajectories of Depressive Symptoms in Middle and Late Life: Racial/Ethnic Variations. Psychol Aging. 2011;26(4):761–77.

41. Hajek A, Brettschneider C, Eisele M, Luehmann D, Mamone S, Wiese B, et al. Disentangling the complex relation of disability and depressive symptoms in old age - findings of a multicenter prospective cohort study in Germany. Int Psychogeriatr. 2017;29(6):885–95.

42. Lee GB, Chang KH, Jae JS. Association between depression and disease-specific treatment. J Affect Disord. 2020;260:124–30.

43. Lei X, Smith JP, Sun X, Zhao Y. Gender Differences in Cognition in China and Reasons for Change over Time: Evidence from CHARLS. J Econ Ageing. 2014;4:46–55.

44. Lei X, Hu Y, McArdle JJ, Smith JP, Zhao Y. Gender Differences in Cognition among Older Adults in China. J Hum Resour. 2012;47(4):951–71.

45. Zhang W, Chen Y, Chen N: Body mass index and trajectories of the cognition among Chinese middle and old-aged adults. BMC Geriatr 2022, 22(1):1.

46. Katz S, Ford AB, Moskowitz RW, Jackson BA, Jaffe MW. Studies of illness in the aged. the index of adl: a standardized measure of biological and psychosocial function. JAMA. 1963;185:914–9.

47. Lawton MP, Brody EM. Assessment of older people: self-maintaining and instrumental activities of daily living. Gerontologist. 1969;9(3):179–86.

48. 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(1):49–57.

49. Fang M, Mirutse G, Guo L, Ma X. Role of socioeconomic status and housing conditions in geriatric depression in rural China: a cross-sectional study. BMJ Open. 2019;9(5):e024046.

50. Suykens JAK, Vandewalle J. Least squares support vector machine classifiers. Neural Proc Lett. 1999;9(3):293–300.

51. Cherkassky V. The nature of statistical learning theory~. IEEE Trans Neural Netw. 1997;8(6):1564.

52. Xuan P, Sun C, Zhang T, Ye Y, Shen T, Dong Y. Gradient boosting decision tree-based method for predicting interactions between target genes and drugs. Front Genet. 2019:10.

53. Byeon H. Developing a random forest classifier for predicting the depression and managing the health of caregivers supporting patients with Alzheimer's Disease. Technol Health Care. 2019;27(5):531–44.

54. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput. 1997;9(8):1735–80.

55. Sapatinas T. The elements of statistical learning. J Royal Stat Soc Series Stat Soc. 2004;167:192.

56. Facal D, Valladares-Rodriguez S, Lojo-Seoane C, Pereiro AX, Anido-Rifon L, Juncos-Rabadan O. Machine learning approaches to studying the role of cognitive reserve in conversion from mild cognitive impairment to dementia. Int J Geriatr Psychiatry. 2019;34(7):941–9.

57. Dinga R, Marquand AF, Veltman DJ, Beekman ATF, Schoevers RA, van Hemert AM, et al. Predicting the naturalistic course of depression from a wide range of clinical, psychological, and biological data: a machine learning approach. Translational. Psychiatry. 2018:8.

58. Gao S, Calhoun VD, Sui J. Machine learning in major depression: From classification to treatment outcome prediction. CNS Neurosci Ther. 2018;24(11):1037–52.

59. Andreescu C, CCH C, Mulsant BH, Ganguli M. Twelve-year depressive symptom trajectories and their predictors in a community sample of older adults. Int Psychogeriatr. 2008;20(2):221–36.

60. Kuo SY, Lin KM, Chen CY, Chuang YL, Chen WJ. Depression trajectories and obesity among the elderly in Taiwan. Psychol Med. 2011;41(8):1665–76.

61. Montagnier D, Dartigues J-F, Rouillon F, Peres K, Falissard B, Onen F. Ageing and trajectories of depressive symptoms in community-dwelling men and women. Int J Geriatr Psychiatry. 2014;29(7):720–9.

62. Fonken LK, Frank MG, Gaudet AD, Maier SF. Stress and aging act through common mechanisms to elicit neuroinflammatory priming. Brain Behav Immun. 2018;73:133–48.

63. Tang M-m, Lin W-j, Pan Y-q, Guan X-t, Li Y-c. Hippocampal neurogenesis dysfunction linked to depressive-like behaviors in a neuroinflammation induced model of depression. Physiol Behav. 2016;161:166–73.

64. Deng P, Gan W, Liu WF, Xie T, Peng GG, Si-Jian LI. The Depression Conditions among Old People in Some Community and the Influential Factors. J Nurs. 2008;15:82.

65. Lu ZQJ. The Elements of Statistical Learning: Data Mining, Inference, and Prediction: World Book Publishing Company; 2008.

Publisher's Note