scientific reports

OPEN

Construction of disability risk prediction model for the elderly based on machine learning

Jing Chen[1,3], Yifei Ren[2,3], Jie Ding[2], Qingqing Hu[2], Jiajia Xu[2], Jun Luo[2], Zhaowen Wu[2] & Ting Chu[2]✉

The study aimed to develop a predictive model using machine learning algorithms, providing healthcare professionals with a novel tool for assessing disability risk in older adults. Data from the 2018 and 2020 waves of the China Health and Retirement Longitudinal Study were utilized, including 3,172 participants aged 65 years and older with no baseline disability. In this study, five machine learning algorithms were employed to construct risk assessment and prediction models for disability in older adults. The Shapley Additive Explanations method was applied to analyze the independent predictors of disability risk. In total, 695 participants (21.9%) were disabled during follow-up. Among the five machine learning models, prediction models constructed using random forest and extreme gradient boosting methods showed superior performance, achieving F1 scores of 0.92 and 0.86 and accuracies of 0.92 and 0.85, respectively. Key predictors of disability risk included self-rated health, education, sleep duration, alcohol consumption, depressive symptoms, hypertension, and arthritis. The Machine learning models for assessing and predicting disability risk in older adults, particularly those developed using RF and XGBoost algorithms, exhibited strong predictive capabilities. These findings highlight the potential of these models for practical application in clinical and public health settings, warranting further exploration and validation.

Keywords Prediction model, Disability, Machine learning, Older adults, Cohort study

Abbreviations

ADL	Activities of daily living
CHARLS	China Health and Retirement Longituitar Study
DCA	Decision curve analysis
DNN	Deep neural network
LR	Logistic regression
ML	Machine learning
PI	Permutation Importance
RF	Random forest.
SHAP	SHapley Additive exPlanations
SVM	Support vector machine
XGBoost	Extreme gradient boosting

Disability is a main problem in the aging process among the population in China and worldwide[1]. Disability refers to limitation or loss of an individual's ability to perform normal activities in daily life, and increases the risk of adverse events such as frailty, falls in the elderly[2]. Disability not only reduces quality of life in older adults, but it also places stress on family caregivers and a burden on the health care system[3]. According to national data, as of January 2020, there are more than 42 million disabled elderly aged 60 and over in China, and the number is expected to reach 77 million by 2030 [4]. Disability is dynamic and progressive. Its occurrence and development are affected by many factors. Therefore, accurately predicting disability risk and developing targeted interventions are conducive to delaying or reducing the adverse effects of disability on quality of life.

To develop an accurate disability prediction model for the elderly, it is essential to first analyze the factors that influence disability. Previous studies have identified several such factors, including age[5], depression[6], smoking

[1]School of Medical Technology and Information Engineering, Zhejiang Chinese Medical University, Hangzhou, Zhejiang, People's Republic of China. [2]School of Nursing, Zhejiang Chinese Medical University, 548 Binwen Road, Binjiang District, Hangzhou 310053, Zhejiang Province, People's Republic of China. [3]Jing Chen and Yifei Ren contributed equally to this work. ✉email: chut@zcmu.edu.cn

history[7], self-rated health[8], sleep duration[9], drinking history[10], hypertension[11] and so on, but the influence of these factors on disability is not completely clear. Most scholars in China and abroad have used logistic regression and Cox regression methods to construct disability risk prediction models for older adults[12,13]. For instance, Chen et al. gathered data from 1,591 elderly individuals in Japan and employed the Cox regression method to construct a model for predicting the risk of disability[14]. The results showed that the area under the receiver operating characteristic (ROC) curve (AUC) of this model was 0.787, and its risk prediction effect was average. Moreover, confounding factors were not excluded in this study, so the applicability of the study findings needs to be verified. At present, there are some studies that use machine learning (ML) algorithm to build a disability prediction model for the elderly[15–17], but these studies have limitations. Wu[17] et al. utilized the CHARLS database and employed a machine learning approach (RF) to analyze predictors of disability among Chinese adults aged 65 and above. The results revealed that walking speed, age, and peak expiratory flow were the most significant predictors. However, this study was restricted to the RF model alone, lacking methodological diversity. In this study, we use five ML algorithms to build a disability prediction model for the elderly, and selects the elderly who have not been disabled at baseline.

ML can reflect the nature of high-dimensional data through supervised, unsupervised, or semi-supervised methods. ML can not only improve the understanding of data but can also be used to analyze high-dimensional, large, and complex relational data. With advances in computing technology during the 21st century, ML has become widely used in the field of health care to help improve the level of disease diagnosis[18–21]. However, ML models are often black boxes, and the visualization of intermediate processes cannot be achieved[22]. Therefore, in this study, the Shapley Additive Explanations (SHAP) and ML algorithms were combined to construct a visual prediction model to explain the individual prediction of kernel-based method[23].

Therefore, the purpose of this study was to establish an accurate and intelligent disability risk assessment and prediction model to provide a new means for the early diagnosis and scientifically based prevention of disability among older adults and help reduce the burden on society owing to disability.

Materials and methods

Study design and participants

The data for this study were from the China Health and Retirement Longitudinal Study (CHARLS). The CHARLS is a nationally representative longitudinal database of people aged 45 years and over in China[24]. The study began a national survey in 2011, covering 150 counties and 450 village/neighborhood committees in 28 provinces of China. The researchers followed participants in 2013, 2015, 2018, and 2020 using face-to-face computer-assisted personal interviewing and gathered extensive information about socioeconomic status, health, and anthropometric and laboratory measurements. The study received ethics approval from the Biomedical Ethics Review Committee of Peking University, Beijing (IRB00001052-11015), and all participants provided informed consent. All study methods were carried out based on the Declaration of Helsinki. All methods were performed in accordance with the relevant guidelines and regulations.

This study used 2018 as the baseline and conducted follow-up in 2020. A total of 19,816 participants were included in the baseline survey, excluding participants aged < 65 years (12,133), those with disability at baseline (2,298), and those with missing or abnormal key predictors (2,213); ultimately, 3,172 participants were included in the analysis.

Disability assessment

The internationally accepted Katz scale was used to evaluate activities of daily living (ADL) of older adults[25]. ADL refers to a person's ability to independently complete necessary activities in daily life, reflecting the most basic ability of self-care[26]. The scale included six items, namely, dressing, bathing, eating, getting in and out of bed, using the bathroom, and controlling urination and defecation. There were four options for each item: (1) no difficulty; (2) difficult but still possible to carry out; (3) difficult and needing help; and (4) unable to carry out. With any one item selected among items (2), (3), or (4), the individual is judged to have disability[14]. Cronbach's α coefficient of the scale in this study was 0.867.

Data collection and participant variables

According to the relevant literature in China and abroad[27–30], Predictors in this study included sociodemographic factors, chronic diseases, health status indicators and lifestyle factors.

Among sociodemographic factors, age was divided into age groups 65–74 years, 75–84 years, and 85 years or above. Marital status comprised married and other, including separated, divorced, widowed, or never married. Education level was categorized as below primary school, primary school, secondary school, or high school and above. The area of residence was divided into rural and urban areas. Among chronic diseases, hypertension, lung disease, arthritis, heart disease, stroke, and diabetes were assessed as presence or absence of each disease. For health status indexes, sleep time was divided into more than 9 h, 6–9 h, and less than 6–9 h. Self-rated health status was divided into very good, good, average, poor, and very poor. Depression was categorized as depression and no depression. Among lifestyle factors, drinking frequency was divided into non-drinking and drinking, smoking frequency into non-smoking and smoking, and exercise into exercising or not exercising every week.

In this study, the 10-item Center of Epidemiologic Studies Depression Scale (CES-D-10) was used to evaluate depression status among participants. The scale comprises 10 items, with a total score of 30. With scores ≥ 10, the patient was classified as having depression; with scores < 10, the patient was classified as not having depression[31].

Derivation and evaluation of prediction models

In this study, the dataset was partitioned using nested cross-validation, in which a super-parametric search is performed by estimating the generalization error of the underlying model to obtain the optimal parameters of the model. Specifically, two loops—the outer loop and inner loop—are included in the nested cross-validation run. The inner layer uses the grid search method to obtain the optimal super-parameters of various models. The outer loop provides 80% of the data as the training set to the inner loop for training while retaining 20% of the data as the test set for testing the inner loop model. In this way, information leakage of data can be prevented to obtain relatively low model scoring deviation.

In this study, we conducted an in-depth investigation of disability in the fields of medicine and nursing, and we pre-screened 18 characteristic variables that are highly related to disability. During our research, we used a least absolute shrinkage and selection operator (Lasso) model to compare the effects, selected 10 additional important characteristic variables, and compared them with the total variable model to verify their functionality. Given the number of feature variables, the correlation, and the size of the training dataset, feature selection was not necessary with sufficient computational resources. Therefore, we did not take feature engineering as the main research focus for the construction of an ML model; we found no serious over-fitting phenomenon in the model.

In this study, nine parameters were used to evaluate the performance of the prediction model, such as accuracy, precision, recall, F1 Score, Hamming loss, Jaccard score, Cohen's kappa score, Confusion matrix and ROC curve and area under curve (AUC). In addition, decision curve analysis (DCA) reflects the prediction performance of the model under different threshold probabilities. All positive and negative lines represent the extreme case of net gain when all samples are positive and negative, respectively. The flowchart for model derivation and verification is shown in Fig. 1.

ML interpretation

As a visualization tool, SHAP can explain ML models by combining the importance of predictors with predictive effects to obtain the relative risk scores of various predictors and then estimating the contribution of each predicted final prediction result from clinical data to predict the probability of the possible occurrence of various clinical events. Because SHAP can be used to quantify the related factors through calculation to improve the accuracy of the prediction results, the performance of the visual prediction model using SHAP is of great importance in improving the accuracy of clinical diagnosis and treatment[32].

Statistical analysis

Continuous variables are presented as the mean ± standard deviation, and categorical variables are presented as frequency and percentage. All the above analyses were conducted using IBM SPSS 27.0. Feature selection, model derivation, and model evaluation were performed with Python 3.7.6. A two-sided p-value of < 0.05 was considered statistically significant. Five distinct machine learning techniques, including logistic regression (LR), deep neural network (DNN), support vector machine (SVM), random forest (RF), and extreme gradient boosting (XGBoost), were chosen to construct models.

Results

Baseline characteristics of the study population

A total of 3,172 participants were enrolled in the study; 695 participants were disabled by the end of follow-up. Sample balancing was performed based on the oversampling method. The sample distribution characteristics of the training and test sets were the same, which indicated that the grouping of the training and verification sets was completely random, and deviation caused by an uneven distribution was avoided. In the comparison of baseline characteristics, there were significant differences in age, marital status, education level, household registration, hypertension, chronic lung disease, arthritis, heart disease, stroke, diabetes, sleep time, depressive symptoms, self-rated health status, smoking, alcohol consumption, and physical exercise. The baseline characteristics of participants are presented in Table 1.

Performance evaluation of prediction models

Results regarding the performance of the predictive model are shown in Table 2. Among comparisons of the five ML algorithms, the prediction model constructed using RF had the best performance. The F1 score and accuracy rate of the model were approximately 0.92; the recall was 0.96; the accuracy rate was 0.89; the Hamming distance was 0.08; the Jaccard coefficient was 0.86; and the kappa coefficient was 0.84. The model constructed using the XGBoost method had good performance, with an F1 score of 0.86, recall of 0.93, and accuracy rate of 0.85. The models constructed using SVM, DNN and logistic regression had average performance. After adding lasso algorithms for feature selection, the F1 scores, recalls, and accuracy rates of the RF and XGBoost models were 0.81 and 0.89; 0.84 and 0.85; and 0.80 and 0.77, respectively. The obtained performance results were relatively poor. This may be owing to the loss of some information during feature selection. The learning curve results showed that there was no serious overfitting of the model, so the results of subsequent presentations were based on the model without LASSO feature selection.

The results of the DCA of the models are shown in Fig. 2A. In general, the performance of several models was consistent with the results in Table 2. It is noteworthy that most models showed a decrease in net benefit with increased threshold value; only the RF model always showed a high net benefit. Based on the observation of its prediction probability results, the model exhibited a very high probability value for the prediction of true positive samples. When the threshold probability was increased to 0.9, the number of false positives was only 1, indicating that RF yielded the optimal clinical value in predicting the normal and abnormal categories. The AUC is shown in Fig. 2B. The results show that the RF model has the highest area under curve (AUC = 0.920). The results of the confusion matrix are shown in Fig. 3. To sum up, these results demonstrate that RF is the optimal model to use.

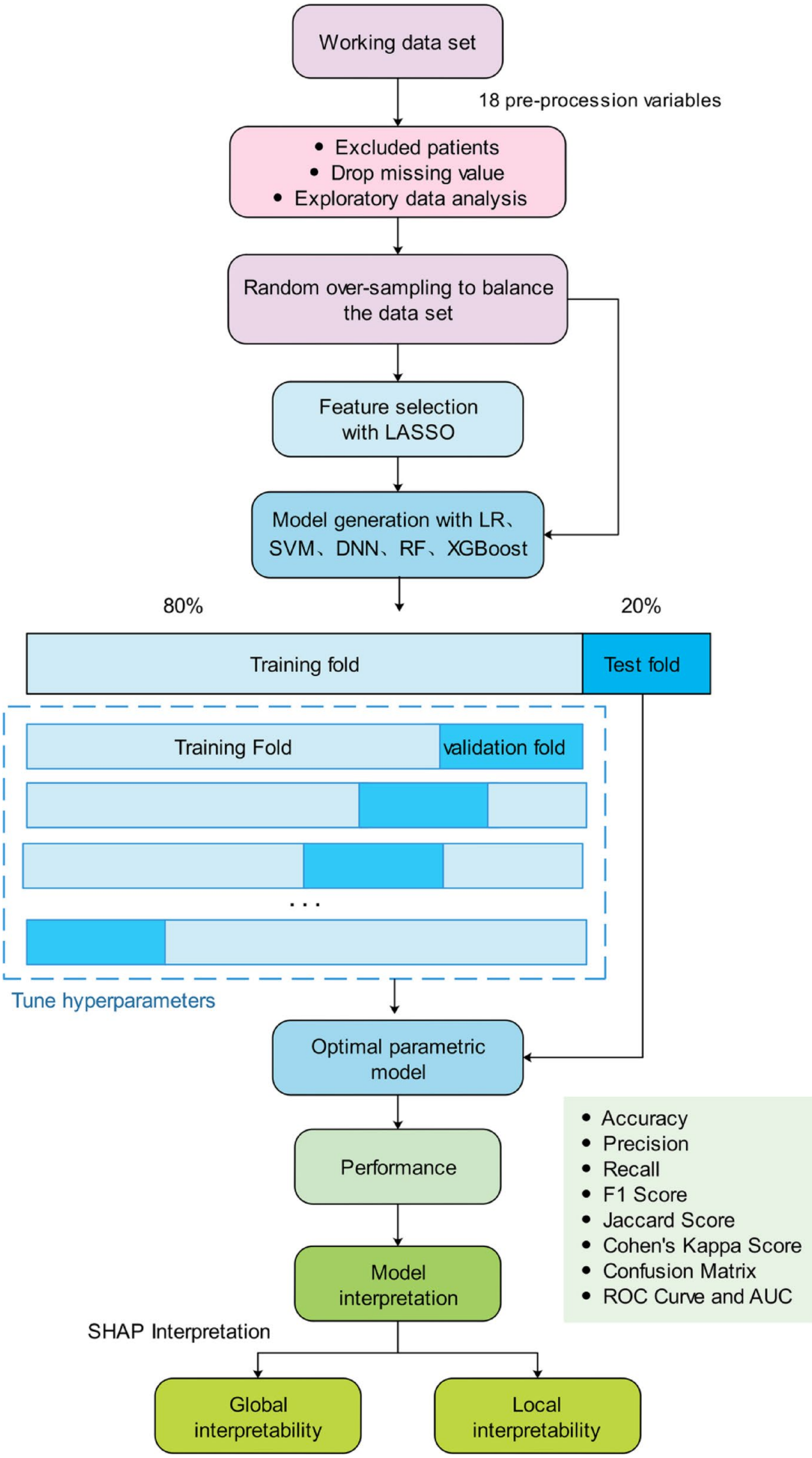

Fig. 1. Flowchart of model derivation and validation.

Interpretability analysis of prediction models

The prediction models constructed using RF, XGBoost, and DNN were visualized using SHAP. The predictors are shown in Fig. 4. The results showed that self-rated health was the most significant predictor of the predictive models constructed using RF. Additionally, sleep duration, educational level, arthritis status, and depressive symptoms all demonstrate high importance with similar rankings in both the Permutation Importance (PI) and

Variables	Overall(4954)	Functional independent (2477)	ADL dependency (2477)	pvalue
Arthritis				< 0.001
No	2580 (52.1)	1508 (60.9)	1072 (43.3)	
Yes	2374 (47.9)	969 (39.1)	1405 (56.7)	
Comorbidity				< 0.001
0	509 (10.3)	335 (13.5)	174 (7.0)	
≥ 1	4445 (89.7)	2142 (86.5)	2303 (93.0)	
Diabetes				< 0.001
No	4062 (82.0)	2080 (84.0)	1982 (80.0)	
Yes	892 (18.0)	397 (16.0)	495 (20.0)	
Drinking				< 0.001
Never	3319 (67.0)	1552 (62.7)	1767 (71.3)	
Drink	1635 (33.0)	925 (37.3)	710 (28.7)	
Education				< 0.001
Below primary school	2506 (50.6)	1093 (44.1)	1413 (57.0)	
Primary school	1284 (25.9)	687 (27.7)	597 (24.1)	
Secondary school	757 (15.3)	437 (17.6)	320 (12.9)	
High school and above	407 (8.2)	260 (10.5)	147 (5.9)	
Exercise				0.004
Never	587 (11.8)	260 (10.5)	327 (13.2)	
Exercise every week	4367 (88.2)	2217 (89.5)	2150 (86.8)	
Self-rated Health				< 0.001
Very poor	318 (6.4)	100 (4.0)	218 (8.8)	
Poor	1306 (26.4)	447 (18.0)	859 (34.7)	
Average	2292 (46.3)	1245 (50.3)	1047 (42.3)	
Good	610 (12.3)	365 (14.7)	245 (9.9)	
Very good	428 (8.6)	320 (12.9)	108 (4.4)	
Heart disease				< 0.001
No	3518 (71.0)	1835 (74.1)	1683 (67.9)	
Yes	1436 (29.0)	642 (25.9)	794 (32.1)	
Hypertension				< 0.001
No	2443 (49.3)	1328 (53.6)	1115 (45.0)	
Yes	2511 (50.7)	1149 (46.4)	1362 (55.0)	
Lung disease				< 0.001
No	3808 (76.9)	1991 (80.4)	1817 (73.4)	
Yes	1146 (23.1)	486 (19.6)	660 (26.6)	
Marital status				0.01
Other	1239 (25.0)	580 (23.4)	659 (26.6)	
Married	3715 (75.0)	1897 (76.6)	1818 (73.4)	
Age (Years)				0.001
65–74	3523 (71.1)	1821 (73.5)	1702 (68.7)	
75–84	1342 (27.1)	615 (24.8)	727 (29.4)	
≥ 85	89 (1.8)	41 (1.7)	48 (1.9)	
Depression				< 0.001
Normal	3073 (62.0)	1752 (70.7)	1321 (53.3)	
Depression	1881 (38.0)	725 (29.3)	1156 (46.7)	
Sleep time (hours)				< 0.001
<6 h	1937 (39.1)	819 (33.1)	1118 (45.1)	
6–9 h	2412 (48.7)	1335 (53.9)	1077 (43.5)	
≥ 9 h	605 (12.2)	323 (13.0)	282 (11.4)	
Residence				< 0.001
Urban	1362 (27.5)	791 (31.9)	571 (23.1)	
Rural	3592 (72.5)	1686 (68.1)	1906 (76.9)	
Smoking				< 0.001
Never	3596 (72.6)	1716 (69.3)	1880 (75.9)	
Smoke	1358 (27.4)	761 (30.7)	597 (24.1)	
Stroke				< 0.001
Continued				

Variables	Overall(4954)	Functional independent (2477)	ADL dependency (2477)	pvalue
No	4442 (89.7)	2264 (91.4)	2178 (87.9)	
Yes	512 (10.3)	213 (8.6)	299 (12.1)	

Table 1. Selected baseline characteristics of participants. Values are presented as number (%).

	Accuracy	Precision	Recall	F1	Haming	Jaccard	Kappa	AUC
LR	0.6640	0.6606	0.6646	0.6626	0.3360	0.4955	0.3279	0.6640
SVM	0.7467	0.7171	0.8089	0.7603	0.2533	0.6133	0.4939	0.7472
DNN	0.7134	0.6694	0.8354	0.7432	0.2866	0.5914	0.4278	0.7143
RF	0.9203	0.8889	0.9593	0.9228	0.0797	0.8566	0.8406	0.9206
Xgboost	0.8526	0.8014	0.9350	0.8630	0.1473	0.7591	0.7057	0.8533

Table 2. Performance evaluation of prediction models. Accuracy, recall, precision, F1 score were all calculated with weighted metrics. Hamming, Jaccard, and Kappa refer to Hamming distance, Jaccard similarity coefficient, and Cohen's kappa score.

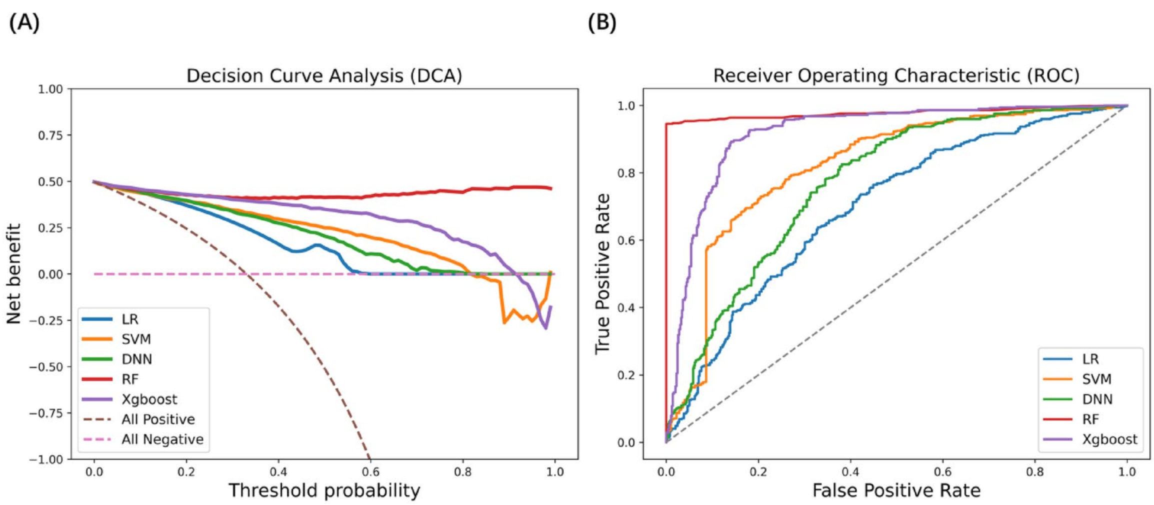

Fig. 2. DCA and AUC for the prediction models.

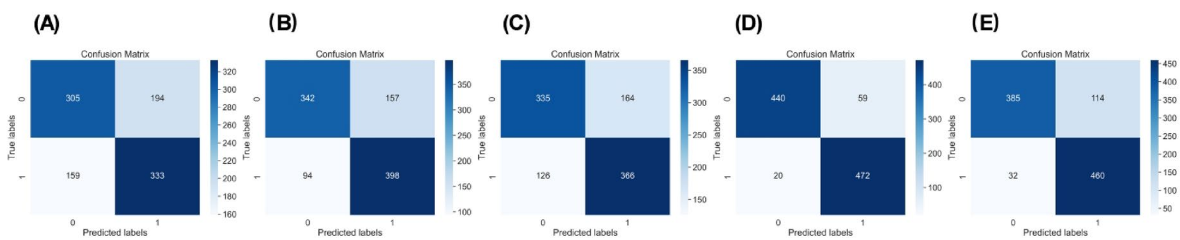

Fig. 3. Confusion matrix. (**A**): Confusion Matrix of LR; (**B**): Confusion Matrix of SVM; (**C**): Confusion Matrix of DNN; (**D**): Confusion Matrix of RF. (**E**): Confusion Matrix of XGBoost.

SHAP analyses. From the perspective of SHAP analysis, the key influencing factors include self-rated health, education level, sleep duration, drinking history, depressive symptoms, hypertension, and arthritis. Although the overall importance distribution trend is consistent, some differences were observed in the ranking of features with moderate to low importance. For instance, the Age variable ranks last in the PI analysis of RF, whereas chronic condition ranks last in SHAP analyses of RF. While age is a known risk factor for disability, its marginal impact on the model's prediction accuracy on the test set is relatively small when the age feature is permuted alone, given that the model can already leverage other strongly correlated variables (such as self-rated health state,

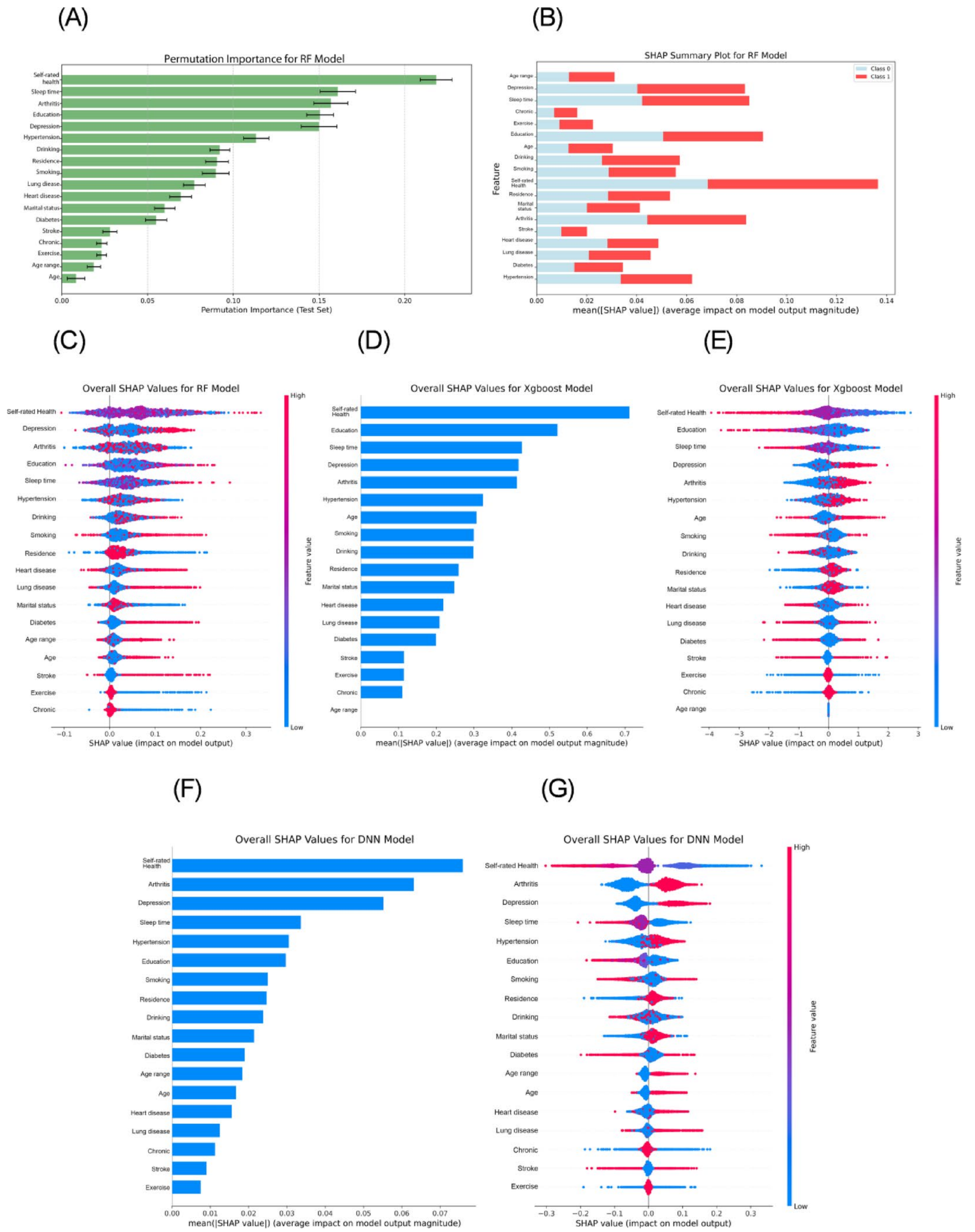

Fig. 4. Relative feature importance of RF (**A**–**C**), XGBoost (**D**–**E**), and DNN (**F**–**G**). Permutation importance of RF; (**B**). SHAP Value of RF; (**C**). SHAP summary of RF; (**D**). Characteristic importance of XGBoost; (**E**). SHAP summary of XGBoost; (**F**). Characteristic importance of DNN; (**G**). SHAP summary of DNN.

arthritis status, etc.) for prediction. Conversely, chronic condition ranks lowest in SHAP, suggesting a smaller role in training fit or average contribution to the prediction magnitude. However, its higher ranking than Age in Permutation Importance implies that removing the information from chronic condition actually impairs the model's predictive performance on the test set slightly more than removing age information. This suggests that chronic condition might retain some unique predictive value for generalization that cannot be fully substituted by age. In the prediction model constructed by XgBoost, self-rated health is the most important predictor, followed by education level and sleep duration. Self-rated health remained the most significant predictor in the DNN-constructed predictive models, followed by arthritis and depressive symptoms. In conclusion, there is a high correlation between self-rated health, education level, sleep duration, drinking history, depressive symptoms, hypertension, and arthritis, and the 2-year disability prediction probability in healthy older adults.

Discussion

In this study, we established a model to predict the 2-year risk of disability risk in older adults aged 65 and above in China who had good health status. We applied five machine learning algorithms to construct the prediction model. Among these models, the random forest model exhibited the best predictive performance. By using SHAP to analyze the best-performing models, we identified key influencing factors, including self-rated health, education level, sleep duration, drinking history, depressive symptoms, hypertension, and arthritis.

We found that self-rated health status was most effective in predicting the disability risk in older adults. Individuals with poor self-rated health often have negative perceptions of their own well-being and may neglect proper health management, which can, in turn, exacerbate disability[33]. Low educational level also increased the risk of disability among older adults[34]. Individuals with low education levels often lack knowledge about health, disease prevention, and chronic condition management, which can increase their risk of disability[35]. Sleep duration is another risk factor for disability in older adults. Chronic sleep deprivation may lead to health issues such as a weakened immune system, metabolic disorders, cardiovascular disease, and hyperglycemia. The onset of these chronic diseases may elevate the risk of disability[9]. Excessive sleep duration often means reduced daily activity, which can lead to issues such as muscle atrophy and osteoporosis, thereby increasing the risk of disability[36]. We found that depression may increase the risk of disability in older adults. This finding is consistent with the results of previous studies[37–39]. Physiologically, depressive symptoms may exacerbate the inflammatory process, thereby increasing the risk of disability. In addition, depression may cause physical fatigue and pain in the elderly, leading to a decline in their ability to perform daily activities[40]. Psychologically, depression in the elderly may lead to reduced medication adherence[41]. We also found that among six chronic diseases, older adults with hypertension or arthritis were at higher risk of disability.

The systematic review[42] indicated that while existing prediction models for disability in older adults demonstrate acceptable discrimination, their overall quality and clinical value remain substantially limited. We acknowledge that current research still exhibits limitations. Based on the findings of the systematic review, our work is somewhat improved. First, data imbalance can affect the performance of the prediction model[43]. Among the subjects, there is an imbalance between disabled and non-disabled elderly individuals. To address this issue, this study employs the oversampling method to mitigate data imbalance, thereby enhancing the prediction accuracy and stability of the model. Regarding data processing, we rigorously excluded samples with pre-existing functional disability at baseline. Secondly, AUC is the most commonly used evaluation metric. In a study that used the same database as this one to predict disability in the elderly, the researchers constructed six models with AUC values ranging from 0.790 to 0.833 [44]. In this study, the AUC of the five machine learning algorithms ranges from 0.664 to 0.920, suggesting that our research can serve as a valuable reference for enhancing disability prediction models for the elderly in China. Additionally, we employed SHAP analysis to improve model interpretability and identify key predictive factors. In the present study, the model performance was evaluated using nested cross-validation and nine parameters, including accuracy, precision, and recall. In addition to the conventional indicators, other indicators were included to comprehensively verify the performance of the model. We found that ensemble learning methods (such as RF and XGBoost) performed best among the five models, which may be because these form a stronger classifier by combining multiple weak learners; thus, if some weak learners make prediction errors, other weak learners can correct the errors to some extent[45]. In summary, by addressing issues such as data imbalance, variable selection, and comprehensive evaluation, we have developed a highly effective disability prediction model for the elderly in China.

There are some limitations in this study. First, the sample size is limited because some subjects are excluded due to missing variables, which may affect the generalization of the model. Second, we can only predict the risk of disability in the elderly two years later. Future studies will need to expand the sample size and increase the prediction period. Thirdly, to improve the universality of the prediction model, external verification should be performed to evaluate the generalization of the model. Finally, in the predictive model in this study, we only considered two states of ADL; disability should be divided into mild, moderate, and severe disability in future studies. Future work should focus on multi-center verification, increasing sample size, and prolonging the time span. This will help to take a substantial step in translating the predictive model from theory to clinical application.

Conclusion

Disability significantly impacts the quality of care and well-being of older adults, imposing a substantial social burden. The results of this study suggest that the use of high-performance ML models for disability risk detection is recommended and has the potential for clinical practice. The ML models developed using RF and XGBoost algorithms demonstrated robust predictive performance and a certain degree of generalizability. These models can effectively identify older adults at high risk of disability, enabling targeted interventions.

ML models offer potential for improving the assessment of disability risk and designing targeted prevention strategies. To mitigate disability risk and promote healthy aging, healthcare professionals can pay more attention to older individuals with low self-rated health scores, low educational levels, short sleep duration, regular alcohol consumption, depressive symptoms, and hypertension or arthritis. Tailored interventions, such as promoting healthy lifestyles and managing blood pressure, might play a pivotal role in preventing disability in these high-risk populations. Future research should focus on external validation using multicenter data to further enhance the applicability and reliability of these models.

Data availability

This research uses data from the China Health and Retirement Longitudinal Study (CHARLS), which can be downloaded at https://charls. charlsdata.com/pages/data/111/zh-cn.html.

Received: 26 January 2025; Accepted: 6 May 2025
Published online: 09 May 2025

References

1. Fried, L. P. & Rowe, J. W. Health in Aging - Past, present, and future. *N. Engl. J. Med.* **383**, 1293–1296. https://doi.org/10.1056/NEJMp2016814 (2020).
2. Steptoe, A., Shankar, A., Demakakos, P. & Wardle, J. Social isolation, loneliness, and all-cause mortality in older men and women. *Proc. Natl. Acad. Sci. USA*. **110**, 5797–5801. https://doi.org/10.1073/pnas.1219686110 (2013).
3. Zhou, J. et al. Development and validation of a nomogram for predicting the 6-Year risk of cognitive impairment among Chinese older adults. *J. Am. Med. Dir. Assoc.* **21**, 864–871e866. https://doi.org/10.1016/j.jamda.2020.03.032 (2020).
4. Luo, Y., Su, B. & Zheng, X. Trends and challenges for population and health during population Aging - China, 2015–2050. *China CDC Wkly.* **3**, 593–598. https://doi.org/10.46234/ccdcw2021.158 (2021).
5. Reichardt, L. A. et al. Unravelling the potential mechanisms behind hospitalization-associated disability in older patients; the Hospital-Associated disability and impact on daily life (Hospital-ADL) cohort study protocol. *BMC Geriatr.* **16**, 59. https://doi.org/10.1186/s12877-016-0232-3 (2016).
6. da Rosa, P. P. S., Marques, L. P., Corrêa, V. P., De Oliveira, C. & Schneider, I. J. C. Is the combination of depression symptoms and Multimorbidity associated with the increase of the prevalence of functional disabilities in Brazilian older adults? A cross-sectional study. *Front. Aging*. **4** https://doi.org/10.3389/fragi.2023.1188552 (2023).
7. Townsend, T. N. & Mehta, N. K. Contributions of obesity and cigarette smoking to incident disability: A longitudinal analysis. *Prev. Med.* **141**, 106226. https://doi.org/10.1016/j.ypmed.2020.106226 (2020).
8. Yamaguchi, R., Makino, K., Katayama, O., Yamagiwa, D. & Shimada, H. Relationship between self-rated health, physical frailty, and incidence of disability among Japanese community-dwelling older adults: A longitudinal prospective cohort study. *Prev. Med.* **191**, 108210. https://doi.org/10.1016/j.ypmed.2024.108210 (2024).
9. Luo, M. et al. Sleep duration and functional disability among Chinese older adults: Cross-sectional study. *JMIR Aging*. **7**, e53548. https://doi.org/10.2196/53548 (2024).
10. Matsuzaki, H. et al. Cut-off point of the risk assessment scale for the 9-year risk of functional disability]. *[Nihon Koshu Eisei Zasshi] Japanese J. Public. Health* **71**, 466–473. https://doi.org/10.11236/jph.23-111 (2024).
11. Wei, R. Enhancing the predictive models for disability in older adults with hypertension: Recommendations for future research. *Psychogeriatrics: Official J. Japanese Psychogeriatr. Soc.* **24**, 1040–1041. https://doi.org/10.1111/psyg.13156 (2024).
12. Sims, T. et al. Simple counts of ADL dependencies do not adequately reflect older adults' preferences toward States of functional impairment. *J. Clin. Epidemiol.* **61**, 1261–1270. https://doi.org/10.1016/j.jclinepi.2008.05.001 (2008).
13. Zhang, L. et al. A web-based dynamic nomogram for predicting instrumental activities of daily living disability in older adults: A nationally representative survey in China. *BMC Geriatr.* **21**, 311. https://doi.org/10.1186/s12877-021-02223-9 (2021).
14. Qi, W., Yin, Z., Sun, Y., Wei, L. & Wu, Y. Nomogram for predicting the 12-year risk of ADL disability among older adults. *Aging Clin. Exp. Res.* **34**, 1583–1591. https://doi.org/10.1007/s40520-022-02105-z (2022).
15. Xiang, C., Wu, Y., Jia, M. & Fang, Y. Machine learning-based prediction of disability risk in geriatric patients with hypertension for different time intervals. *Arch. Gerontol. Geriatr.* **105**, 104835. https://doi.org/10.1016/j.archger.2022.104835 (2023).
16. Speiser, J. L. et al. Predicting future mobility limitation in older adults: A machine learning analysis of health ABC study data. *J. Gerontol. A* **77**, 1072–1078. https://doi.org/10.1093/gerona/glab269 (2022).
17. Wu, Y. et al. Examining individual and contextual predictors of disability in Chinese older adults: A machine learning approach. *Int. J. Med. Informatics*. **191**, 105552. https://doi.org/10.1016/j.ijmedinf.2024.105552 (2024).
18. Nooripour, R. et al. How resiliency and hope can predict stress of Covid-19 by mediating role of spiritual Well-being based on machine learning. *J. Relig. Health*. **60**, 2306–2321. https://doi.org/10.1007/s10943-020-01151-z (2021).
19. Senoner, T. et al. Identifying the location of an accessory pathway in Pre-Excitation syndromes using an artificial Intelligence-Based algorithm. *J. Clin. Med.* **10** https://doi.org/10.3390/jcm10194394 (2021).
20. Sawano, S. et al. Deep learning model to detect significant aortic regurgitation using electrocardiography. *J. Cardiol.* **79**, 334–341. https://doi.org/10.1016/j.jjcc.2021.08.029 (2022).
21. Liu, J. et al. Clinical parameters and metabolomic biomarkers that predict inhospital outcomes in patients with ST-Segment elevated myocardial infarctions. *Front. Physiol.* **12**, 820240. https://doi.org/10.3389/fphys.2021.820240 (2021).
22. Zihni, E. et al. Opening the black box of artificial intelligence for clinical decision support: A study predicting stroke outcome. *PLoS One*. **15**, e0231166. https://doi.org/10.1371/journal.pone.0231166 (2020).
23. Parsa, A. B., Movahedi, A., Taghipour, H., Derrible, S. & Mohammadian, A. K. Toward safer highways, application of XGBoost and SHAP for real-time accident detection and feature analysis. *Accid. Anal. Prev.* **136**, 105405. https://doi.org/10.1016/j.aap.2019.105405 (2020).
24. Zhao, Y., Hu, Y., Smith, J. P., Strauss, J. & Yang, G. Cohort profile: the China health and retirement longitudinal study (CHARLS). *Int. J. Epidemiol.* **43**, 61–68. https://doi.org/10.1093/ije/dys203 (2014).
25. Katz, S. et al. A standardized measure of biological and psychosocial function. *Jama* **185**, 914–919. https://doi.org/10.1001/jama.1963.03060120024016 (1963).
26. Katz, S. et al. *J. Gerontol.* **19**, 285–293, doi:https://doi.org/10.1093/geronj/19.3.285 (1964).
27. Cheng, S. et al. Trajectories and influencing factors of cognitive function and physical disability in Chinese older people. *Front. Public. Health* **12**, 1380657. https://doi.org/10.3389/fpubh.2024.1380657 (2024).
28. Wu, Y., Xiang, C., Jia, M. & Fang, Y. Correction: interpretable classifiers for prediction of disability trajectories using a nationwide longitudinal database. *BMC Geriatr.* **23**, 190. https://doi.org/10.1186/s12877-023-03843-z (2023).
29. Lu, Y. et al. Machine Learning-Based prediction of functional disability: A cohort study of Japanese older adults in 2013–2019. *J. Gen. Intern. Med.* **38**, 2486–2493. https://doi.org/10.1007/s11606-023-08215-2 (2023).
30. Lu, S. et al. Prevalence and possible factors of cognitive frailty in the elderly with hypertension and diabetes. *Front. Cardiovasc. Med.* **9**, 1054208. https://doi.org/10.3389/fcvm.2022.1054208 (2022).
31. Andresen, E. M., Malmgren, J. A., Carter, W. B. & Patrick, D. L. Screening for depression in well older adults: Evaluation of a short form of the CES-D (Center for epidemiologic studies depression Scale). *Am. J. Prev. Med.* **10**, 77–84 (1994).
32. Deshmukh, F. & Merchant, S. S. Explainable machine learning model for predicting GI bleed mortality in the intensive care unit. *Am. J. Gastroenterol.* **115**, 1657–1668. https://doi.org/10.14309/ajg.0000000000000632 (2020).
33. Enroth, L. & Fors, S. Trends in the social class inequalities in disability and Self-Rated health: Repeated Cross-Sectional surveys from Finland and Sweden 2001–2018. *Int. J. Public. Health* **66**, 645513. https://doi.org/10.3389/ijph.2021.645513 (2021).
34. Carmona-Torres, J. M. et al. Disability for basic and instrumental activities of daily living in older individuals. *PloS One* **14**, e0220157. https://doi.org/10.1371/journal.pone.0220157 (2019).
35. Sun, F., Zimmer, Z. & Zajacova, A. Pain and disability transitions among older Americans: The role of education. *J. Pain*. **24**, 1009–1019. https://doi.org/10.1016/j.jpain.2023.01.014 (2023).
36. Peng, Y. et al. Association of nocturnal sleep duration and nocturnal sleep changes with instrumental activities of daily living disability among middle-aged and elderly Chinese. *Sleep Med.* **109**, 90–97. https://doi.org/10.1016/j.sleep.2023.06.017 (2023).
37. Dong, L., Freedman, V. A. & de Mendes, C. F. The association of comorbid depression and anxiety symptoms with disability onset in older adults. *Psychosom. Med.* **82**, 158–164. https://doi.org/10.1097/psy.0000000000000763 (2020).

38. Wang, J. et al. Exploring the reciprocal relationship between activities of daily living disability and depressive symptoms among middle-aged and older Chinese people: A four-wave, cross-lagged model. *BMC Public. Health.* **23**, 1180. https://doi.org/10.1186/s12889-023-16100-0 (2023).
39. Yu, Q. et al. Predicting functional dependency using machine learning among a middle-aged and older Chinese population. *Arch. Gerontol. Geriatr.* **115** https://doi.org/10.1016/j.archger.2023.105124 (2023).
40. Morin, R. T., Nelson, C., Bickford, D., Insel, P. S. & Mackin, R. S. Somatic and anxiety symptoms of depression are associated with disability in late life depression. *Aging Ment. Health* **24**, 1225–1228. https://doi.org/10.1080/13607863.2019.1597013 (2020).
41. Kong, D., Solomon, P. & Dong, X. Depressive symptoms and onset of functional disability over 2 years: A prospective cohort study. *J. Am. Geriatr. Soc.* **67**, S538–s544. https://doi.org/10.1111/jgs.15801 (2019).
42. Zhou, J. et al. Risk prediction models for disability in older adults: A systematic review and critical appraisal. *BMC Geriatr.* **24**, 806. https://doi.org/10.1186/s12877-024-05409-z (2024).
43. Wang, L. et al. Classifying 2-year recurrence in patients with Dlbcl using clinical variables with imbalanced data and machine learning methods. *Comput. Methods Programs Biomed.* **196**, 105567. https://doi.org/10.1016/j.cmpb.2020.105567 (2020).
44. Han, Y. & Wang, S. Disability risk prediction model based on machine learning among Chinese healthy older adults: Results from the China health and retirement longitudinal study. *Front. Public. Health* **11**, 1271595. https://doi.org/10.3389/fpubh.2023.1271595 (2023).
45. Motwani, M. et al. Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease: A 5-year multicentre prospective registry analysis. *Eur. Heart J.* **38**, 500–507. https://doi.org/10.1093/eurheartj/ehw188 (2017).

Acknowledgements
We would like to thank the China Health and Retirement Longitudinal Study (CHARLS) for providing us with the data. The sponsors had no role in the design, methods, subject recruitment, data collection, analysis, and preparation of this paper.

Author contributions
Study concept and design: Chen, Ren, Ding, Chu; Analysis and interpretation of data: Chen, Ren, Xu, Chu; Drafting of the manuscript: Chen, Ren, Hu, Chu; Critical revision of the manuscript for important intellectual content: Chen, Ren, Luo, Wu, Chu. All authors read and approved the final manuscript.

Funding
This work is supported by the Research Project of the Zhejiang Provincial Medical and Health Science and Technology Plan grant number 2024KY127.

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to T.C.