scientific reports

OPEN

Enhancing the convenience of frailty index assessment for elderly Chinese people with machine learning methods

Li Huang[1,2,3], Huajian Chen[2,3] & Zhenzhen Liang[1✉]

Frailty is a state that is closely associated with adverse health outcomes in the aging process. The frailty index (FI), which measures frailty in terms of cumulative deficits, has been widely used for frailty assessment in elderly people, and its advantage of self-reported information collection makes it applicable to a broader group of elderly people. Our study aims to simplify the Frailty Index Assessment Scale, while maintaining its reliability and accuracy, to easily and quickly assess frailty in elderly people. In this study, participants (age ≥ 65 years) from the Chinese Longitudinal Healthy Longevity Survey (CLHLS), which had 13,339, 372 and 1214 participants in 2008, 2011, and 2014, respectively, were used. The 2008 dataset was split into 80% for training and 20% for internal validation, and the data from 2011 to 2014 as external validation. In order to obtain effective predictors, we used Lasso regression, Boruta algorithm and random forest classifier score for feature selection. We used six models for predictive model construction and evaluated the models in the validation dataset. Model performance was measured by area under the curve (AUC), accuracy and F1 score. Logistic regression was found to be the best performing and most interpretable algorithm with AUC, accuracy and F1 of 0.974, 0.932 and 0.880 for the validation dataset, respectively. The AUCs for the external independent validation dataset were 0.963 and 0.977, respectively. Subgroup analysis showed that the model had good predictive power in both males and females. The predictive power was stronger among the elderly people over 80 years old, with AUC, accuracy and F1 of 0.973,0.914, and 0.893, respectively. The model also obtained good predictive power in the case of FI measured by different indicators. The model showed good robustness in the follow-up assessment of frailty status in elderly people, with the AUC remaining above 0.95 and accuracy above 0.9 over the long-term follow-up. Using machine learning techniques, we have successfully developed a simple frailty assessment prediction model based on 10 key features to shorten the frailty assessment scale with near full-scale accuracy. A user-friendly website was created to facilitate the application of this prediction model (https://healthy-aging.shinyapps.io/Frailty_Assessment/).

Keywords Elderly people, Machine learning, Frailty index, Shortened questionnaire

Abbreviations
CLHLS	Chinese Longitudinal Healthy Longevity Survey
FI	Frailty Index
LR	Logistic Regression
RF	Random Forest
SHLNN	Single-hidden-layer Neural Network
SVM	Support Vector Machine
XGBoost	Extreme gradient boosting
AUC	Area under the curve

Frailty reflects the physical and psychological health deficits and social deficits that accumulate during the aging process, as well as the degradation of the organism's defenses due to loss of physiological reserves[1–3], a state that

[1]School of Public Health, Xinxiang Medical University, Xinxiang 453003, China. [2]School of Public Health, Wenzhou Medical University, Wenzhou 325035, China. [3]Li Huang and Huajian Chen contributed equally to this work and share first authorship. ✉email: 221010@xxmu.edu.cn

is strongly associated with adverse health outcomes, including disability, dependency, falls, need for long-term care, and death[4–6]. And this state usually appears and develops in elderly people[4,5], with a higher prevalence of frailty in those over 80 years of age. The current situation of rapidly aging populations is a particularly serious challenge in global public health, which has led to an increase in the number of frail elderly people and further burdened the current public health care system. Frailty is considered to be an early stage of disability. And, because it is reversible, this suggests that appropriate intervention at the right time can prevent, delay or even reverse this state[7].

The frailty index (FI), which calculates the proportion of individual deficits, can be used to characterize the frailty status of elderly people, as suggested by Rockwood et al.[3] and Mitnitsk et al.[8]. And the frailty index describing frailty status has been well validated and widely used in the assessment of frailty in elderly people, and applies to a broader population of elderly people than the well-known definition of frailty proposed by Fried et al.[9] (a syndrome or phenotype of at least three of the five criteria: weight loss, fatigue, weak grip, slow walking speed, and low physical activity), and as EAMONN et al.[10] reported, older inpatients are often unable to complete performance-based tests. The number of indicators used to construct the FI in previous studies using FI to measure frailty varies, for example, Goggins et al.[11] used 62 indicators, Gu et al.[12] used 39 indicators, Clegg et al.[5] used 20 indicators and Sha et al.[13] used 37 indicators. These indicators used to construct the frailty index all have the common feature of having a large number of items (≥ 20 items). Despite strong predictive validity, even if participants are willing to answer all questions regularly, repeated use of the same questions is likely to lead to burnout and memory responses rather than responses based on currently experienced symptoms.

Recent research advances show that machine learning can provide effective mitigation for these problems[14,15]. Sun et al. used machine learning to shorten diagnostic scales to assess a person's depressive state without sacrificing accuracy[16]. Wang et al.[17] assessed cognitive impairment in elderly people using 4,6 and 9 varying features and machine learning techniques and their best model achieved an AUC of 0.8269. Recently a study from Spain reported that the FRAIL scale needs to be obtained through interviews and cannot be obtained directly from electronic medical records. To conveniently assess the frailty status of elderly people through electronic medical records, researchers constructed a frailty prediction model based on easy-to-collect variables with good model performance, and developed a mobile app[18]. However, there is still less research on the diagnostic assessment of frailty status in the field of machine learning, especially for the frailty index. Constructing simple and efficient predictive models for assessment is still of high research value. Especially in China, a country with a large population, the aging population is increasingly aggravated[19], the complexities within the elderly people population and the large base size of each group require a widely applicable, convenient and reliable frailty assessment scale to quickly and easily screen frail elderly people and provide appropriate interventions.

The model in this study was designed to predict frailty status based on FI, a derived indicator calculated by integrating multiple variables to assess an individual's frailty status. Therefore, based on the national survey of elderly people in 2008, 2011 and 2014, this study used machine learning algorithms to build a widely applicable and simple frailty risk assessment prediction model based on the FI to identify the occurrence of frailty in elderly people.

Materials and methods
Data sources
The data were obtained from the Chinese Longitudinal Health and Longevity Survey (CLHLS) organized by Peking University, which was conducted in 23 provinces of China among people aged 65 and older and their adult children aged 35–64. This survey, which began in 1998 and is examined every 2–3 years, is the most extensive and longest-running social science survey in China. About half of the cities and counties in the 23 research provinces were randomly selected as research sites for the survey. CLHLS systematically collects data on elderly people through face-to-face interviews conducted by trained staff[20]. The data quality of the CLHLS was validated as acceptable[21], as well as more detailed descriptions (sampling design, response rate, and questionnaire validity) were published widely elsewhere[22,23]. The CLHLS study was approved by the Research Ethics Committee of Peking University (IRB00001052-13074), and all participants or their surrogate respondents provided a written informed consent.

Our study used the 2008, 2011 and 2014 CLHLS datasets. A detailed flow chart of participant selection is shown in Supplementary Figure S1. Of the 16,954 respondents interviewed in 2008, 13,339 respondents aged 65–105 years (one study suggested that the quality of declarations may be poorer for ages 106 and older in the survey[24]) completed estimates of frailty indices as well as complete demographic characteristics. Of the 9765 respondents in 2011, 5027 respondents aged 65–105 years completed estimates on the frailty index. Of the 7192 respondents in 2014, 2482 respondents aged 65–105 years completed estimates on the frailty index. Immediately afterward, we matched participants from between the three datasets. In the 2011 and 2014 data, we removed duplicate interviewed participants (2008 participants were removed from the 2011 data and 2008 and 2011 participants were removed from the 2014 data). Finally, 14,925 participants were included in this study, with 13,339, 372, and 1,214 participants in 2008, 2011, and 2014, respectively.

Definition of frailty
Frailty status as assessed by the frailty index (FI) was used as a predictor variable for this study. Following the standard procedure recommended by Searle and co-authors[25], our study used 44 frailty indicators to form the FI[26], including aspects of self-care, cognitive functioning, medical conditions, self-rated health, and mental health status. Each frailty indicator was dichotomized or multichotomized and then mapped to a 0–1 interval (e.g., visual function, where "can see and distinguish" was assigned 0, "can see only" was assigned 0.33, "can not see" was assigned 0.67, and "blind" was assigned 1) to reflect its severity. For each participant, FI was equal to the sum of the scores on the 44 frailty indicators and divided by the theoretical maximum score of 45. FI is a

continuous variable ranging from 0 to 1, with higher FI scores indicating greater participant frailty. Based on previous studies reported[6,13,27], we turned the continuous frailty index into a dichotomous variable: non-frail (≤ 0.21) and frail (FI > 0.21). A full description of the frailty index can be found in Supplementary Table S1.

Demographic characteristics
We also included demographic characteristics as candidate predictors, including age (continuous variable, categorical variable: 65–79 vs. 80–105 years), sex (male vs. female), marital status (married vs. other), place of residence (city, town and rural), education (formal vs. informal by years of education ≥ 1 and < 1), occupation (occupation before age 60 classified as "agricultural work" and "non-agricultural work"), current alcohol consumption (yes or no), current smoking (yes or no), current exercise (yes or no), and sleep duration (sleep duration was classified into three categories: ≤ 6 h, 7–9 h, and ≥ 10 h).

Feature selection
Feature selection allows extracting key features as a way to reduce the complexity of the model without losing too much total information. Three methods were used for feature selection: Least absolute shrinkage and selection operator (Lasso) regression[28], Boruta[29] and random forest classifier scoring[30]. Lasso is an efficient feature selection and regularization method. It filters out smaller and more relevant subsets of predictor variables from a potentially multicollinear set of variables by shrinking the regression coefficients. This method is particularly effective when dealing with high-dimensional data, as it enables variable selection and reduces the risk of overfitting. Boruta algorithm is an advanced feature selection method based on random forest (RF), which identifies all the features related to the dependent variable by creating "shadow features" and comparing their importance. It calculates feature importance scores through multiple iterations and marks features that exceed the maximum importance of "shadow features" as relevant. This method introduces additional randomness to reduce the influence of misleading correlation, improves the stability and reliability of feature selection, and is especially suitable for dealing with complex datasets, and can effectively identify truly important features. Random forest classifier scoring is a feature selection method based on RF that uses the intrinsic properties of the RF model to assess the importance of features, which can effectively handle high-dimensional data, capture nonlinear relationships, and be insensitive to outliers. Lasso is suitable for linear relationships, while Boruta and RF can capture non-linear relationships. This combination improved the robustness of the feature selection and reduced the possible bias of a single method.

Construction of prediction model
The 2008 data with 13,339 participants were divided into 80% as the training dataset and 20% as the validation dataset. Feature filtering was performed on the training dataset to obtain the important features. During parameter tuning of each model, a grid search and 10-fold cross-validation are used to find the best performing parameters in the training dataset. Five machine learning algorithms were trained, including logistic regression (LR), RF, support vector machine (SVM), extreme gradient boosting (XGBoost), and single-hidden-layer neural network (SHLNN). In this study, we adopted the Stacking approach to integrate five base models (LR, RF, SVM, XGBoost, and SHLNN). Stacking is an integrated learning strategy that builds a hierarchical prediction framework by combining the prediction results of different individual learners. The method contains at least two levels of learners, and is therefore also known as a hierarchical model integration framework. In this study, we implemented a two-level stacked framework. First, predictions were generated using five trained base models and these results were used as new feature inputs. Subsequently, we employed the LR algorithm as a meta-learner to train these new features to construct the final prediction model. With this approach, we effectively integrated the prediction results from each base model, which may help improve the overall predictive performance of the model. The six models used in this study covered the spectrum from simple linear to complex nonlinear, and by systematically comparing models of different complexity, we were able to comprehensively assess the prediction performance and sought the optimal balance between model complexity and interpretability[17,31]. The model evaluation was performed using AUC, accuracy and F1 score[16] in the validation dataset (Fig. 1). Using external independent validation datasets from 2011 to 2014, the top-performing models were revalidated for model validity and generalizability. AUC is a composite measure of the algorithm's ability to distinguish between outcome classes at all possible classification thresholds, ranging from 0.5 to 1. Accuracy is a metric of a classification model that measures the number of correct predictions as a percentage of the total number of predictions made, ranging from 0 to 1. The F1 score is a measure of the accuracy of a binary classification model, taking into account both the accuracy and recall of the classification model, ranging from 0 to 1. Higher AUC, accuracy and F1 score indicate better prediction performance.

Using the simplified model to evaluate FI measured by different indicators
Our goal is to validate that the simplified model of FI we developed is also robust in FI with different metric measures. For this purpose, we collected the FI designed by Gu et al.[12] in 2009 and four FI published in recent years (Lv et al.[32], Chen et al.[33], Xu et al.[34] and Zhang et al.[35]). In separate datasets from 2011 to 2014, we used these five FI calculations to measure participants' frailty status and then evaluated model performance using ROC-AUC values.

Using the simplified model to track and assess elderly people from 2008 to 2014
We intended to confirm that the model has strong predictive power in tracking participants' frailty because CLHLS data are characterized by follow-up. Therefore, we chose 2008 as the starting point and evaluated participants for the 2011 and 2014 follow-ups to ensure that we could obtain a sufficiently long follow-up period and a sufficient number of participants to, in turn, more accurately assess the predictive power of the
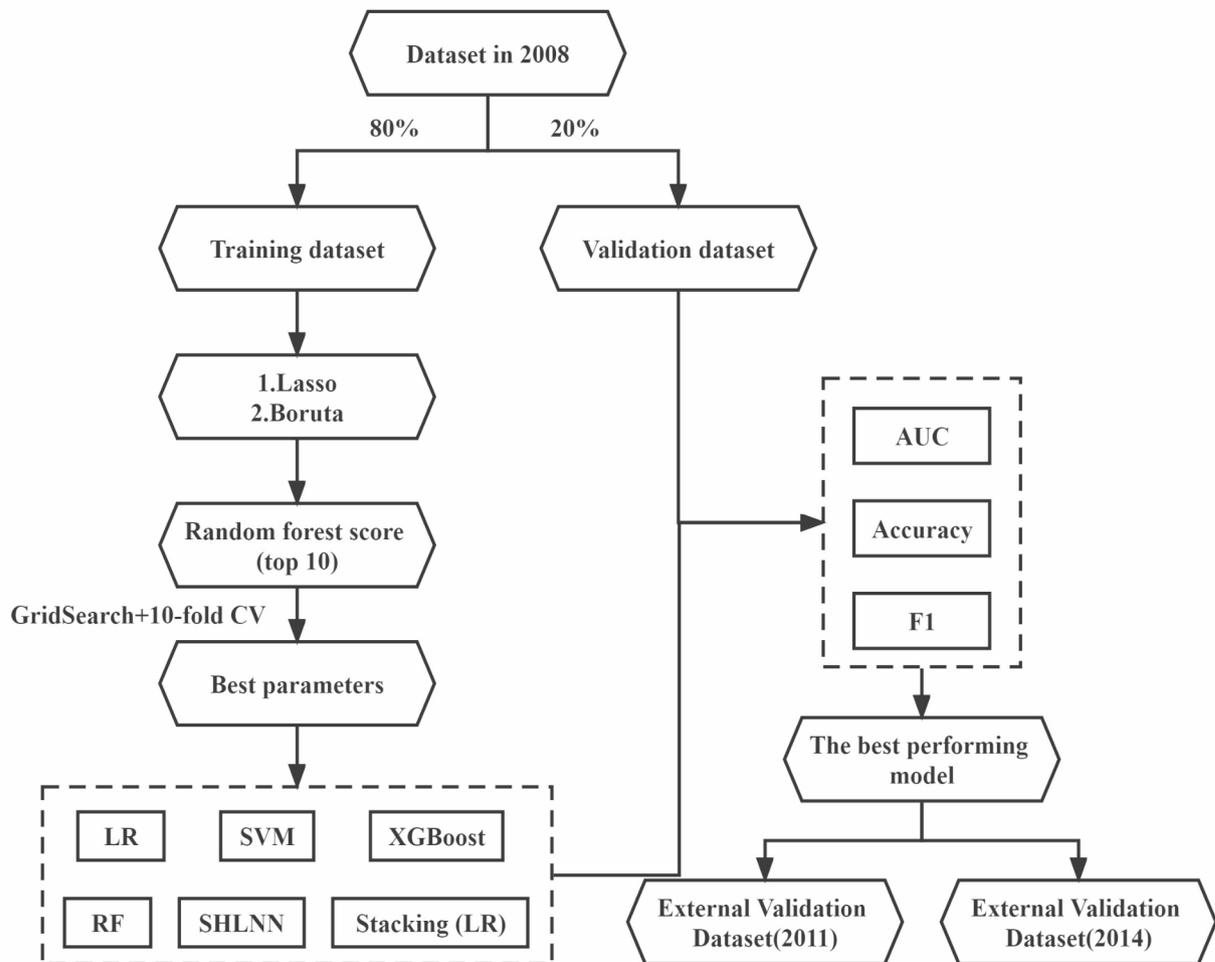
Fig. 1. Schematic diagram of our prediction framework.

model. ROC-AUC values and accuracy were utilized as metrics to assess the predictive power of the model. The screening of the elderly people follow-up assessment sample is shown in Fig. 2.

Construction and application of the frailty assessment website
To enable rapid application of the model, we have built a frailty assessment website using the shiny package for R and the "Shinyapps.io" program for Rstudio to apply and transform the model. Shinyapps.io can host shiny applications, it is easy to use, secure and scalable and requires no hardware.

Statistical analysis
Categorical variables were expressed as counts (percentages) and continuous variables were expressed as means with standard deviations (SD). We directly excluded samples with missing values in the frailty indicators and demographic characteristics. The frailty index recommended by Searle and co-authors was used, including 44 frailty indicators. The histogram was used to depict the change in frequency of the frailty index in 2008, 2011 and 2014. The model was interpreted using binomial logistic regression and a forest plot was drawn. Subgroup analysis of the best model was performed using gender and age grouping. All statistical analyses were performed using R (4.1.2) software. The R packages used include glmnet, Boruta, randomForest, e1071, xgboost, nnet, pROC, caret, tidyverse and forestplot. A p-value < 0.05 was deemed statistically significant.

Results
Distribution of frailty index
The histogram of the frailty index in Supplementary Figure S2 showed the distribution of the frailty index of the participants in 2008,2011 and 2014, and the results showed that the mean frailty index in 2008, 2011 and 2014 was 0.17,0.23 and 0.16, respectively, as well as the mean frailty index for these three years was 0.17.
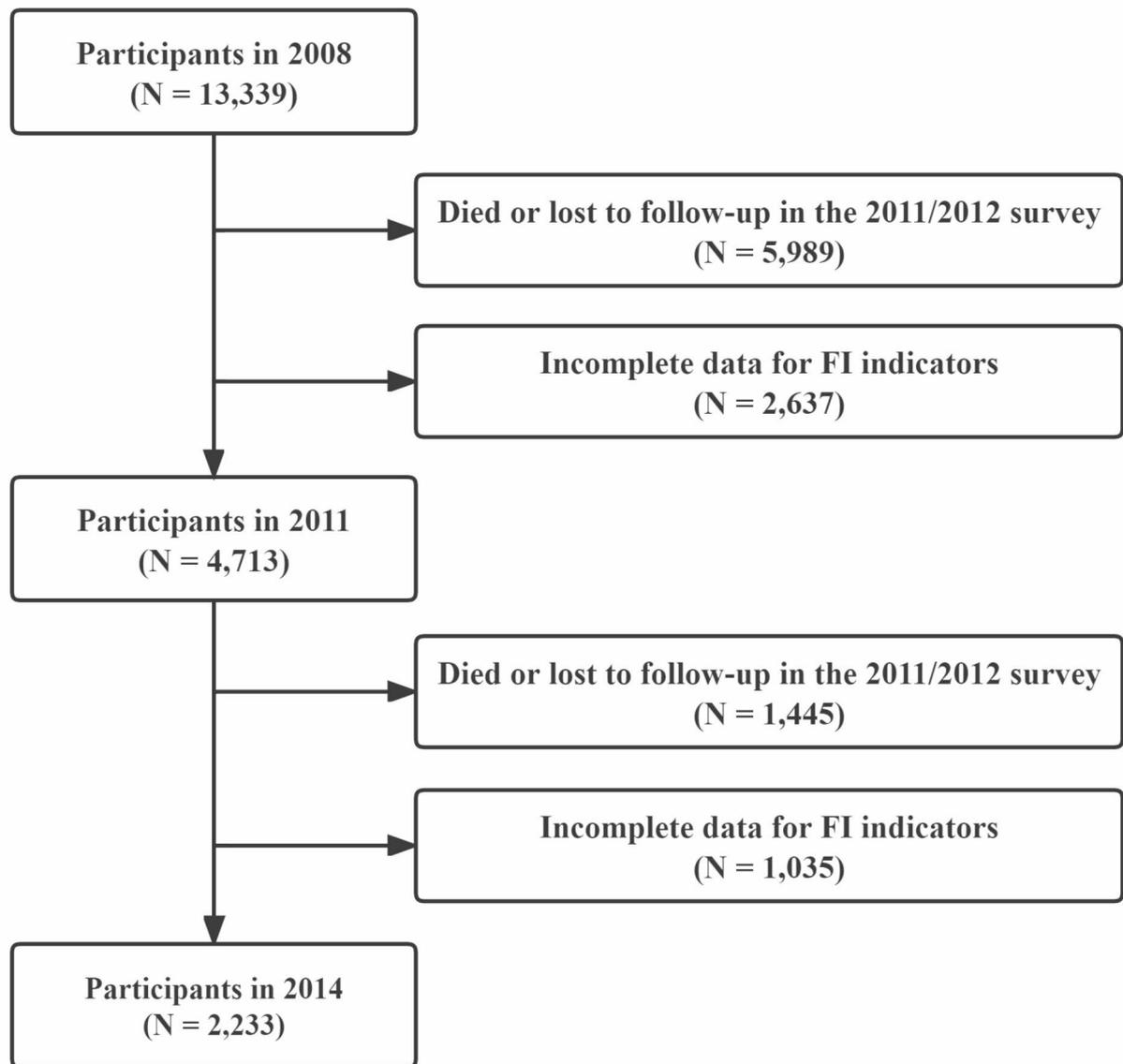
Participants in 2008
(N = 13,339)

Died or lost to follow-up in the 2011/2012 survey
(N = 5,989)

Incomplete data for FI indicators
(N = 2,637)

Participants in 2011
(N = 4,713)

Died or lost to follow-up in the 2011/2012 survey
(N = 1,445)

Incomplete data for FI indicators
(N = 1,035)

Participants in 2014
(N = 2,233)

Fig. 2. Flow chart of sample screening for follow-up assessment of elderly people.

Dataset segmentation and demographic characteristics

We utilized the training dataset to train and build the model, the validation dataset to verify the performance of the model, and the external validation datasets from 2011 to 2014 to again verify the validity and generalization ability of the model. Supplementary Table S2 showed the characteristics of all datasets. There were 10,672 participants in the training dataset with a mean age of 85.3 ± 10.9, of which 45.9% (4895) were male and 54.1% (5777) were female, and 28.9% (3088) were frail. In the validation dataset there were 2667 participants with a mean age of 85.3 ± 11.1, 45.2% (1206) males and 54.8% (1461) females, 28.9% (771) frail. There were 372 and 1214 participants in the 2011 and 2014 datasets, respectively, of which the mean age was 89.5 ± 11.8 and 85.4 ± 11.0, respectively, with 43.5% and 46.9% of males, 56.5% and 53.1% of females, 51.3% and 27.4% of frailty.

Feature selection

First, we performed a 10-fold cross-validation Lasso regression, selecting a lambda with one standard error from the minimum of Binomial Deviance for coefficient compression (Fig. 3A-B and Supplementary Table S3), and we obtained 50 candidate features. We also utilized the Boruta algorithm for feature filtering and selected 49 candidate features with Confirmed. Then, we intersected the candidate features obtained by Lasso regression and the Boruta algorithm (Fig. 3C), and obtained 45 candidate features common to both methods. Finally, the 45 candidate features were incorporated into the random forest classifier using the Gini coefficient to rate the
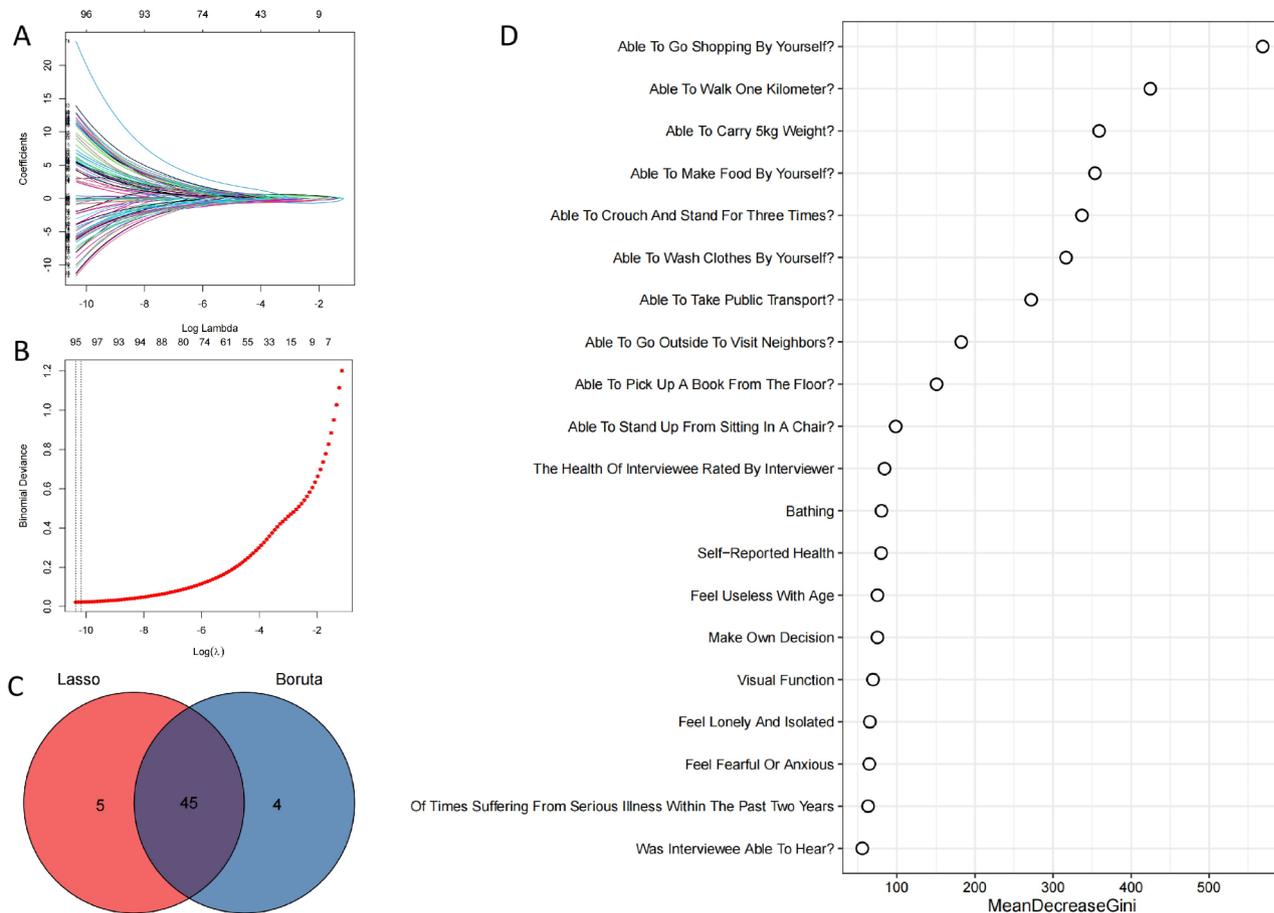
Fig. 3. Feature selection. (**A**) The lasso regression curve. (**B**) The 10-fold cross-validation parameter (λ) options. (**C**) Venn diagram of common features selected by Lasso and Boruta. (**D**) MeanDecreaseGini score of random forest classifier.

Model	AUC	Accuracy	F1
Logistic Regression	0.974	0.932	0.880
Random Forest	0.967	0.928	0.870
Support Vector Machine	0.974	0.930	0.868
XGBoost	0.974	0.932	0.880
Single-hidden-layer Neural Network	0.974	0.930	0.874
Fusion (Stacking)	0.974	0.933	0.881

Table 1. Performance of machine learning models in the validation dataset.

importance of the features, and the top 10 key features were selected in descending order of rating (Fig. 3D). The distribution of the 10 key features is described in Supplementary Table S2.

Prediction models
Results of the algorithm performance are shown in Table 1 and the best parameters for models are shown in Supplementary Table S4. Having the exception of RF, all of the models' ROC curves in the validation dataset displayed high predictive power, with an AUC of 0.974. For accuracy, LR and XGBoost have the best performance with an accuracy of 0.932 for both. In the F1 score, LR and XGBoost also showed the best performance, with F1 score of 0.880. After fusing all five models, the results improved the accuracy to 0.933 and the F1 score to 0.881, although the AUC was not improved.

External validation of the best model
Due to the similar predictive performance of the models, we finally chose the more easily interpretable LR as the frailty risk simple assessment prediction model. To clarify the connection between each trait and the risk of frailty, multivariate analysis using binomial logistic regression was carried out, and a forest plot was created

(Supplementary Figure S3). The results showed significant associations between all 10 characteristics and frailty risk, as well as all being independent predictors of frailty risk. Then, the model robustness was again validated using the 2011 and 2014 external independent validation sets, which showed an AUC of 0.963 (Fig. 4A) for the 2011 external independent validation dataset, as well as 0.977 (Fig. 4B) for the 2014 external independent validation dataset. Regarding accuracy, we have 0.884 and 0.933 in the 2011 and 2014 datasets, respectively (Supplementary Table S5). And F1 scores of 0.885 and 0.878 in the 2011 and 2014 datasets, respectively (Supplementary Table S5). Two external validation datasets are used to test the model performance once again, and the results show that the model is both robust and capable of generalization.

Subgroup analysis

We conducted research to determine which demographic category the model has the best predictive ability. Utilizing AUC, Accuracy, and F1 scores Again to gauge the model's strength in terms of predictive potential, we conducted subgroup analyses with gender and age groups. By combining the validation dataset with the 2011 and 2014 external independent datasets into separate data, we were able to analyze subgroups. The results showed (Supplementary Table S6) that the AUC and accuracy of the model showed good predictive ability for both gender and age groupings, with AUCs greater than 0.95 and accuracy greater than 0.9. And when we assessed the model's performance using the F1 score, we discovered that, with the exception of the 65–79 age group, where the F1 score was 0.698, the F1 score was greater than 0.850 in both the male and female and 80–105 age groups. The outcomes of the aforementioned subgroup analysis show that the model is capable of making accurate predictions for both males and females. However, the model's predictive evaluation in people aged 65 to 79 was poor, but it performed best in those aged 80 to 105 (AUC = 0.973, Accuracy = 0.914, F1 = 0.893).

Evaluation of FI calculated for different indicators

We evaluated the FI constructed with five different indicators using the simplified model. The results showed that the AUC values of 0.934 for Gu et al. and Lv et al., 0.900 for Chen et al., 0.875 for Xu et al. and 0.901 for Zhang et al. showed strong predictive power in the external independent validation data in 2011 (Fig. 5A). In the 2014 external independent validation data, Gu et al. and Lv et al. showed similarly strong predictive power with AUC values of 0.952, Chen et al. with 0.900, Xu et al. with 0.889, and Zhang et al. with 0.920 (Fig. 5B).

Using the model to track and evaluate participants from 2008 to 2014

Through the follow-up characteristics of the CLHLS data, we used the model to follow up and assess the frailty status of elderly people (Fig. 6). For the 2008 data with 13,339 elderly people, the AUC and accuracy were 0.966 and 0.924, respectively, and for the 2011 data with 4,713 elderly people, the AUC and accuracy were 0.956 and 0.923, respectively. For the 2014 data with 2,233 elderly people, the AUC and accuracy were 0.972 and 0.917, respectively. All of these follow-up assessments demonstrated that the simplified model also has good predictive power in tracking frailty in elderly people.
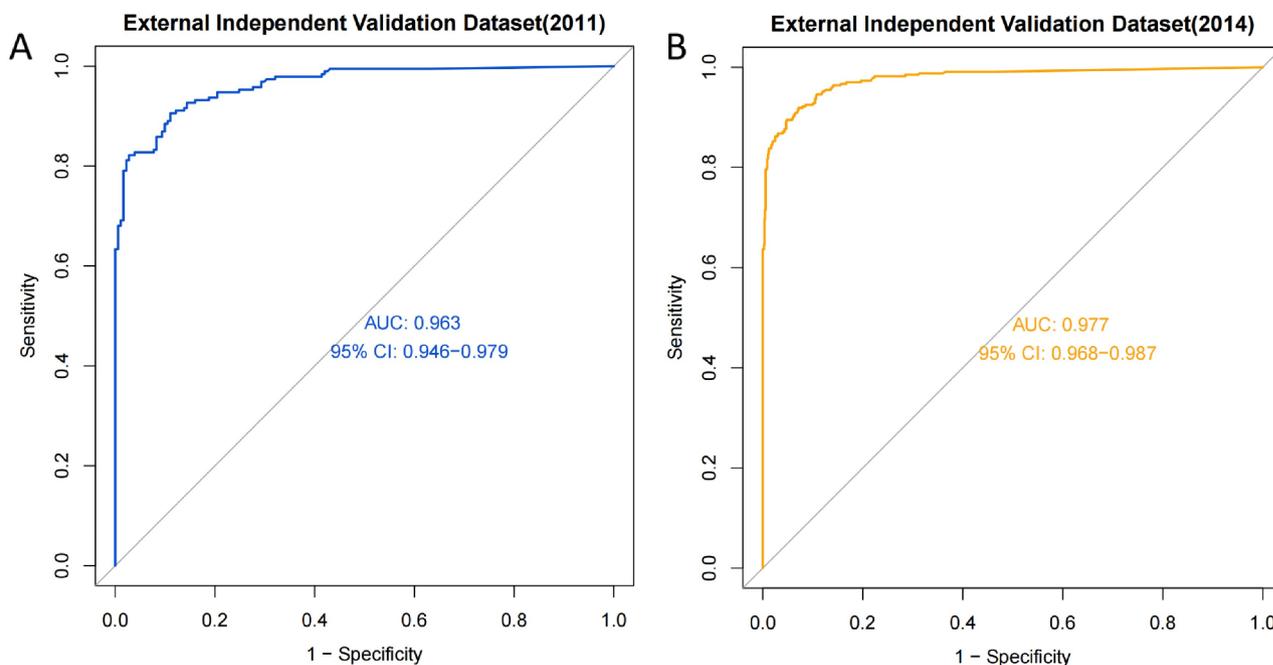

Fig. 4. ROC curves and AUCs for 2011 (**A**) and 2014 (**B**) external independent validation datasets.
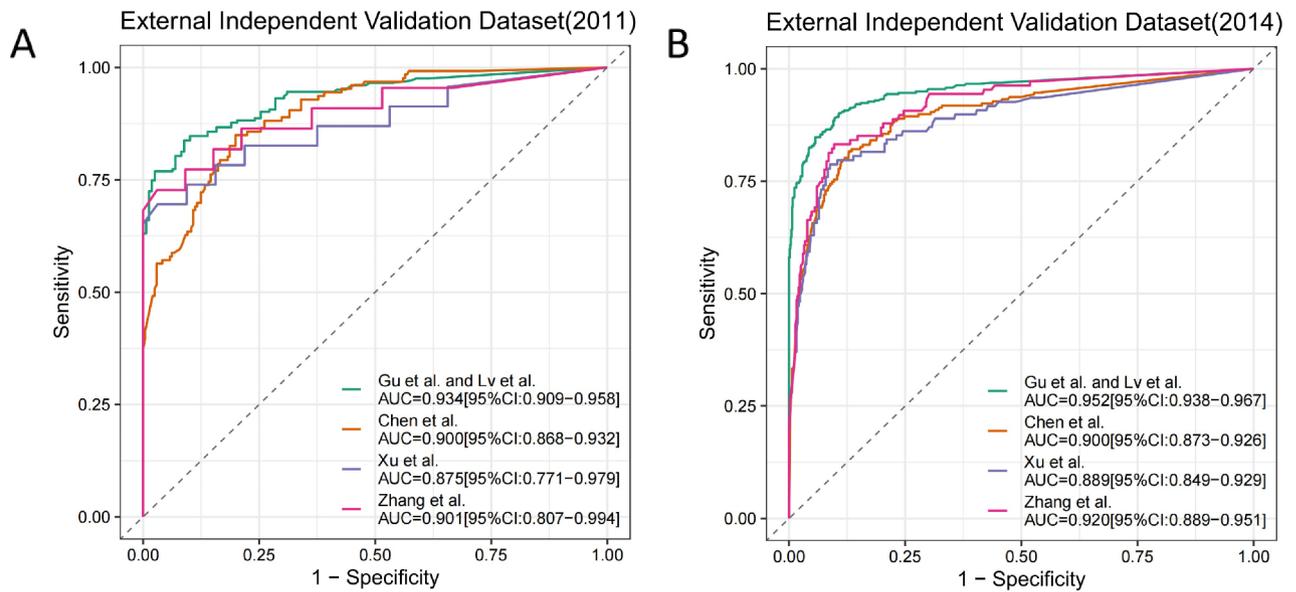
Fig. 5. ROC curves and AUC of FI constructed by the model for five different independent studies, including two external independent validation sets 2011 (**A**) and 2014 (**B**).
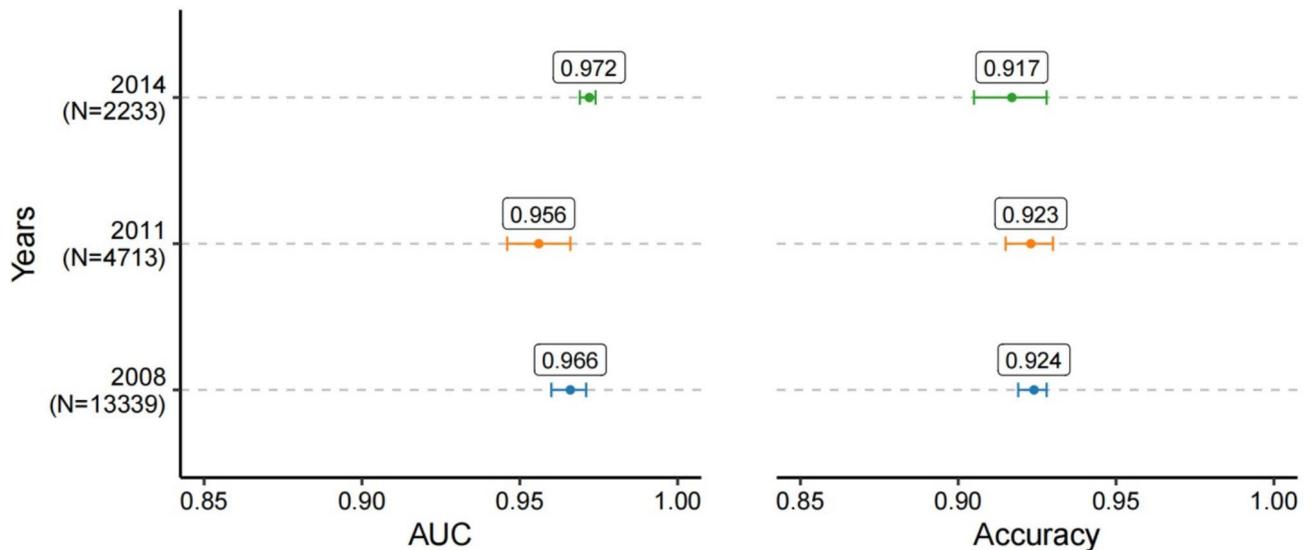

Fig. 6. Model follow-up assessed the AUC and accuracy of the follow-up data for 2008, 2011 and 2014.

Frailty assessment website

To make it easier for others to use our model, we have created a user-friendly website (https://healthy-aging.shinyapps.io/Frailty_Assessment/). The website contains 10 key features for frailty prediction, and after the 10 key features are selected (Fig. 7A), click "Predict", and the graph of frailty probability and non-frailty probability will appear in the first row of the "Prediction" column on the right (Fig. 7B). The graph of frailty probability and non-frailty probability will appear in the "Prediction" column on the right (Fig. 7B), and the data will be recorded and saved in real-time in the second row of the "New Observations" column (Fig. 7C).

Discussion

Our research has successfully simplified the computation of the frailty index by applying multiple machine learning methods to ensure that short and long questionnaires have similar accuracy in assessing frailty, and developed a convenient and easy-to-use web-based tool for frailty prediction, which makes the assessment of frailty more efficient and convenient. Frailty is a syndrome associated with multiple chronic conditions and multiple risk factors[36,37], including psychosocial and functional limitations. It is also a precursor state of functional limitation and disability associated with the individual aging process itself. Frailty is also an important
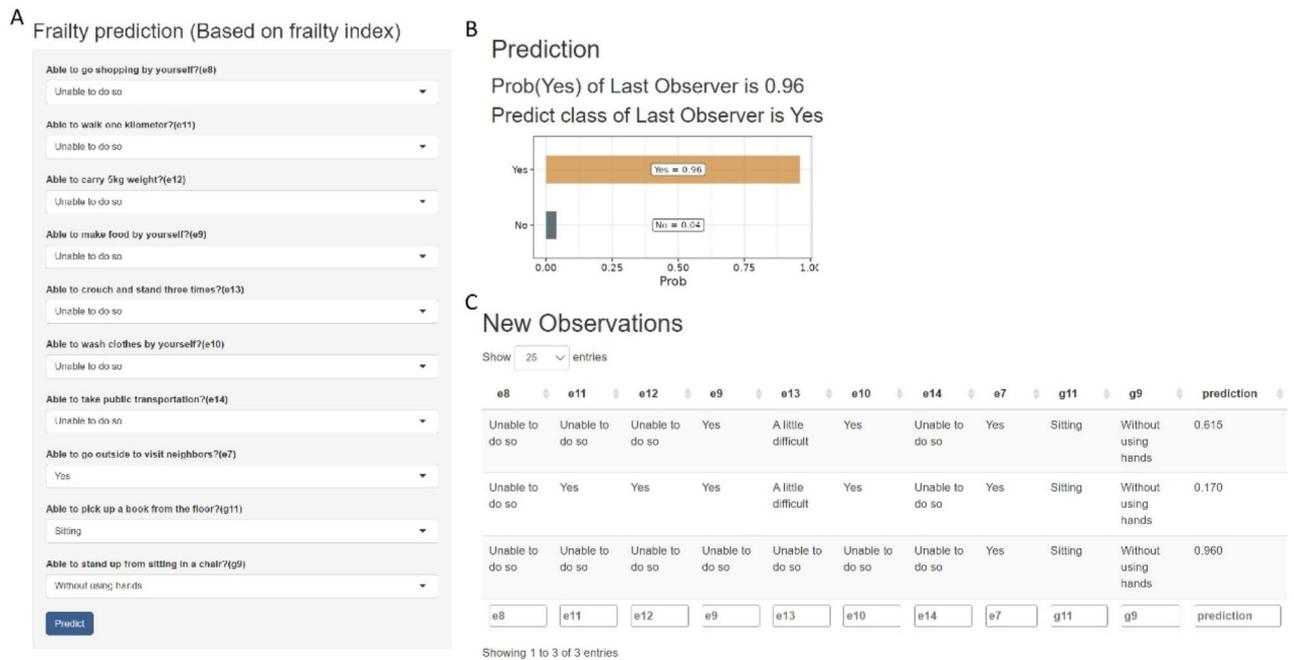
Fig. 7. The Web page for frailty assessment, including the feature options column (**A**), the predicted frailty probability column (**B**), and the results saving column (**C**).

concept in the study of morbidity and mortality in old age. Because frailty is a dynamic and changeable state, prompt assessment and action can prevent it or even reverse it, safeguarding the health and quality of life of elderly people.

FI is a comprehensive health index that combines subjective and objective indicators to provide a more comprehensive picture of the true health status of elderly people. The degree of frailty is quantified by cumulative health loss, which is a symptom of the illness and health of the organism and reflects the health deficit of the physical and conscious features of the life course[38]. The FI is widely used in international health studies of elderly people, with a large number of studies in the fields of social gerontology and epidemiology in particular[39–41]. This important indicator has become a useful tool for geriatricians, clinicians and other practitioners in public health surveillance and intervention. Therefore, studying the FI of our elderly people is of great social importance to achieve healthy aging.

To assess individual frailty status with a similar level of accuracy as the long questionnaire, this study applied a machine learning technique to derive a shorter questionnaire from a longer one. To achieve this goal, we chose a frailty index scale containing 44 frailty indicators developed by Gao et al.[26] in Chinese elderly people according to the standard procedure recommended by Searle and co-authors[25]. We used this scale as a basis for obtaining a shorter version to assess the likelihood of predicting a person's frailty status. The results of our study largely confirm our hypothesis.

To begin with, we hypothesized that a multiple-to-few-question approach would be able to select only a few items from the full set of frailty indicator scale items to predict an individual's frailty status without affecting the accuracy of the original questionnaire. In fact, we obtained the top 10 features by incorporating Lasso and Boruta's intersection features into the random forest classifier for feature scoring, and these features were much more important than the features after 10, and most importantly, the accuracy of the model based on 10 features reached more than 90% of the original questionnaire and the F1 score reached more than 0.85.

We further hypothesize that following feature selection, different machine learning algorithms exhibit varying prediction performance and accuracy under the critical features. Therefore, we considered both advanced prediction models and traditional prediction models. We discovered that traditional logistic regression gave results that were just as good as those produced by RF, SVM, XGBoost, and SHLNN with more advanced algorithms. Due to LR's great interpretability, it is better suited for the prediction of frailty status. In addition, we confirmed the hypothesis that a fusion model (stacking) based on multiple machine learning models should produce the best results, as it can pool the strengths of multiple models. However, the gain from using the fusion model is small, with an accuracy increase of only 0.003.

The scores for the features' Gini coefficients showed an inflection point at the tenth place in the order from largest to smallest, which was where we intercepted the first ten features. Among the top ten characteristics, eight of them belong to questions that assess instrumental daily living ability and two belong to questions that assess physical health. Here we have a question, age is the main factor of aging, why is age not selected as a key characteristic? Our reading through the literature revealed that actual age is readily accepted as being associated with frailty. However, as argued elsewhere[3,37], the concept of frailty is rooted in the fact that actual age as an explanatory model is imprecise when predicting adverse outcomes in individual cases. The frailty index itself

serves as a proxy for biological age and is a better representation of aging in elderly people than actual age[12]. Because frailty is independent of actual age, it provides unique insight into the occurrence of adverse health outcomes from aging-related processes[40].

Although the shortened assessments we developed cannot completely replace the original scales, they are considerably shorter and easier to use than the original scales, especially in China, which has a large population base. When we predict the likelihood that a respondent has a frailty status, we can effectively guide the older person through a comprehensive assessment as well as timely interventions for the older person. Therefore, the tool we developed will make it easier and faster to screen for frailty status. On the staff side, this will help reduce workload and increase motivation to complete the questionnaire; on the elderly people surveyed side, this will help reduce questionnaire boredom and reduce the risk of disability and death; and on the individual, family, and societal side, this will help reduce physical, mental, and economic burdens.

We conducted an exploration of the predictive power of the model in the subgroup. It's interesting to note that the model performs well in both men and women. In terms of age group, the prediction ability of the model is the best in the age group of 80–105 years, with an F1 of 0.893, while the prediction ability is not satisfactory in the age group of 65–79 years. We hypothesized that since the elderly people in the CLHLS data were mainly concentrated around the age of 85, this was also evidenced by the average age of participants in the training dataset of 85.3 years. The age groups of 65–79 years and 80–105 years in the training dataset accounted for 30.88% and 69.12%, respectively, and the imbalance between age groups resulted in the model's predictive power being biased toward the 80–105 years age group. The American Medical Association reports that 40% of elderly people aged ≥ 80 years have frailty[42], suggesting that the model is better able to help screen and intervene at the appropriate time in the high-incidence age group.

We aim to demonstrate that the model remains well robust and reliable under the FI of other indicators. We explored the relationship between the FI measured by these different indicators and the model using 1 classical FI measure established by Gu et al. and 4 recent FI measured by different indicators. The model obtained good performance, all with good AUC values, indicating that the simplified model based on 10 key features can be applied to FI measured by different indicators, and that the model is well adapted to FI.

It is also critical for the model to track and evaluate the effects. The accuracy of long-term forecast results is more meaningful than short-term forecasts. By tracking and evaluating over time, a more accurate assessment of the model's performance and effectiveness on long-term time scales can be obtained. Therefore, we further determined the persistence and reliability of the model by evaluating data from three waves of followers in 2008, 2011, and 2014, and the results similarly demonstrated the good performance of the model, indicating that the model can be well adapted for follow-up assessment.

Finally, we built a frailty assessment webpage based on the frailty index in order to translate the research results of this study for interactive application with reality, which facilitates a fast as well as easy frailty risk assessment and can be used as an aid to quickly screen out frail elderly people.

Our study's advantages included national representation, a sizable sample size, and participant validation across time. We anticipate that this study will provide estimates of frailty status based on current health status, particularly in elderly people in the high frailty prevalence age group of 80 years and older. However, our study has some limitations. (1) Some factors in this study are from self-report, and there may be some subjectivity of the elderly people, which may lead to information bias. However, self-reported information is readily available in preventive care. (2) There is no fixed quantitative criterion for the number of frailty indicators in the frailty index scale, and we have selected only one frailty index scale for questionnaire shortening, which may have problems in application. But, to test whether the level and distribution of frailty among Canadian seniors would change, Mitnitski et al.[39] constructed a frailty index using various numbers of variables. They discovered that the frailty index did not change significantly depending on the number of variables used, indicating that the frailty index was robust and not sensitive to the number of variables used to construct it. (3) This study focuses specifically on participants aged 65 years and over in China. Considering the differences in cultural and genetic backgrounds of this particular population, the results of the study may not be applicable to other racial/ethnic populations. (4) The model in this study did not directly assess the predictive performance on the frailty phenotype (weight loss, fatigue, reduced grip strength, etc.), but only for FI, aiming to reduce the number of features required for FI and simplify the calculations, but future studies should further validate the potential of the model to be applied to the frailty phenotype.

While our current study focused on frailty assessment, feature selection and machine learning methods can be used to estimate other important clinical metrics associated with aging, such as cognitive function, cardiovascular risk, or functional impairment. By selecting relevant features and applying similar modeling techniques, it is possible to predict outcomes in these domains with high accuracy, providing a comprehensive risk assessment tool for elderly people. In addition, future research can be based on the model framework of this study to answer a series of new scientific questions, such as whether other health indicators have similar predictive patterns? Can the methodology of this study also be used for simplification in health assessments that involve a large amount of information? This will provide a more comprehensive scientific basis for health management of the elderly people.

Conclusions
In our study, we used machine learning techniques, and successfully developed a simple frailty assessment prediction model based on 10 key features to shorten the frailty assessment scale with near full-scale accuracy. A user-friendly website was created to facilitate the application of this prediction model (https://healthy-aging.shinyapps.io/Frailty_Assessment/). This study can effectively identify elderly people with frailty so that implement intervention early to optimize the allocation of healthcare resources and raise public awareness of elderly health issues. Meanwhile, our study provides new perspectives on frailty status assessment, enriches the

academic literature in the field of geriatrics, and provides new insights for future research on adverse health outcomes.

Data availability

References

1. Collard, R. M., Boter, H., Schoevers, R. A. & Oude Voshaar, R. C. Prevalence of frailty in community-dwelling older persons: A systematic review. *J. Am. Geriatr. Soc.* **60**(8), 1487–1492 (2012).
2. Rockwood, K. et al. A global clinical measure of fitness and frailty in elderly people. *Cmaj* **173**(5), 489–495 (2005).
3. Rockwood, K., Hogan, D. B. & MacKnight, C. Conceptualisation and measurement of frailty in elderly people. *Drugs Aging* **17**(4), 295–302 (2000).
4. Fried, L. P., Ferrucci, L., Darer, J., Williamson, J. D. & Anderson, G. Untangling the concepts of disability, frailty, and Comorbidity: Implications for improved targeting and care. *J. Gerontol. Ser. A* **59**(3), M255–M63 (2004).
5. Clegg, A., Young, J., Iliffe, S., Rikkert, M. O. & Rockwood, K. Frailty in elderly people. *Lancet* **381**(9868), 752–762 (2013).
6. Zhu, A., Yan, L., Wu, C. & Ji, J. S. Residential greenness and Frailty among older adults: A longitudinal cohort in China. *J. Am. Med. Dir. Assoc.* **21**(6), 759–65e2 (2020).
7. Lorenzo-López, L. et al. Nutritional determinants of frailty in older adults: A systematic review. *BMC Geriatr.* **17**(1), 108 (2017).
8. Mitnitski, A. B., Song, X. & Rockwood, K. The estimation of relative fitness and frailty in community-dwelling older adults using self-report data. *J. Gerontol. Biol. Sci. Med. Sci.* **59**(6), M627–M632 (2004).
9. Fried, L. P. et al. Frailty in older adults: Evidence for a phenotype. *J. Gerontol. Biol. Sci. Med. Sci.* **56**(3), M146–M156 (2001).
10. Eeles, E. M., White, S. V., O'Mahony, S. M., Bayer, A. J. & Hubbard, R. E. The impact of frailty and delirium on mortality in older inpatients. *Age Ageing* **41**(3), 412–416 (2012).
11. Goggins, W. B., Woo, J., Sham, A. & Ho, S. C. Frailty index as a measure of biological age in a Chinese population. *J. Gerontol. Biol. Sci. Med. Sci.* **60**(8), 1046–1051 (2005).
12. Gu, D. et al. Frailty and mortality among Chinese at advanced ages. *J. Gerontol. B Psychol. Sci. Soc. Sci.* **64**(2), 279–289 (2009).
13. Sha, S., Xu, Y. & Chen, L. Loneliness as a risk factor for frailty transition among older Chinese people. *BMC Geriatr.* **20**(1), 300 (2020).
14. Chikersal, P. et al. Detecting depression and predicting its onset using longitudinal symptoms captured by passive sensing: A machine learning approach with robust feature selection. *ACM Trans. Comput. Hum. Interact.* **28**(1), Article 3 (2021).
15. Qiao, J. & editor A systematic review of machine learning approaches for mental disorder prediction on social media. In *2020 International Conference on Computing and Data Science (CDS)*, 1–2 Aug. 2020 (2020).
16. Sun, Y. H., Liu, Q., Lee, N. Y., Li, X. & Lee, K. A novel machine learning approach to shorten depression risk assessment for convenient uses. *J. Affect. Disord.* **312**, 275–291 (2022).
17. Wang, S. et al. Using machine learning algorithms for predicting cognitive impairment and identifying modifiable factors among Chinese elderly people. *Front. Aging Neurosci.* **14**, 977034 (2022).
18. Aznar-Tortonda, V. et al. Detection of frailty in older patients using a mobile app: Cross-sectional observational study in primary care. *Br. J. Gen. Pract.* **70**(690), e29–e35 (2020).
19. Man, W., Wang, S. & Yang, H. Exploring the spatial-temporal distribution and evolution of population aging and social-economic indicators in China. *BMC Public Health* **21**(1), 966 (2021).
20. Shen, K., Zhang, B. & Feng, Q. Association between tea consumption and depressive symptom among Chinese older adults. *BMC Geriatr.* **19**(1), 246 (2019).
21. Feng, Y., Liu, E., Yue, Z., Zhang, Q. & Han, T. The evolutionary trends of Health behaviors in Chinese Elderly and the influencing factors of these trends: 2005–2014. *Int. J. Environ. Res. Public. Health* **16**(10) (2019).
22. Zeng, Y. Towards deeper research and better policy for healthy aging—Using the unique data of Chinese longitudinal healthy longevity survey. *China Econ. J.* **5**(2–3), 131–149 (2012).
23. Yi, Z., Vaupel, J. W., Zhenyu, X., Chunyuan, Z. & Yuzhi, L. The Healthy Longevity Survey and the Active Life Expectancy of the Oldest Old in China. Population: An English Selection. ;13(1):95–116. (2001).
24. Yi, Z. *Reliability of Age Reporting among the Chinese Oldest-Old in the CLHLS Datasets* (Springer Netherlands).
25. Searle, S. D., Mitnitski, A., Gahbauer, E. A., Gill, T. M. & Rockwood, K. A standard procedure for creating a frailty index. *BMC Geriatr.* **8**, 24 (2008).
26. Gao, T. et al. Long-term tea consumption reduces the risk of frailty in older Chinese people: Result from a 6-year longitudinal study. *Front. Nutr.* **9**, 916791 (2022).
27. Hoover, M., Rotermann, M., Sanmartin, C. & Bernier, J. Validation of an index to estimate the prevalence of frailty among community-dwelling seniors. *Health Rep.* **24**(9), 10–17 (2013).
28. McEligot, A. J., Poynor, V., Sharma, R. & Panangadan, A. Logistic LASSO regression for dietary intakes and breast Cancer. *Nutrients* **12**(9) (2020).
29. Kursa, M. B. & Rudnicki, W. R. Feature selection with the Boruta Package. *J. Stat. Softw.* **36**(11), 1–13 (2010).
30. Rahman, M. S., Rahman, M. K., Kaykobad, M., Rahman, M. S. & isGPT An optimized model to identify sub-golgi protein types using SVM and Random Forest based feature selection. *Artif. Intell. Med.* **84**, 90–100 (2018).
31. Wu, Y., Xiang, C., Jia, M. & Fang, Y. Interpretable classifiers for prediction of disability trajectories using a nationwide longitudinal database. *BMC Geriatr.* **22**(1), 627 (2022).
32. Lv, Y. et al. Long-term fine particular exposure and incidence of frailty in older adults: Findings from the Chinese longitudinal healthy longevity survey. *Age Ageing* **52**(2) (2023).
33. Chen, Q. et al. Dynamic statistical model for predicting the risk of death among older Chinese people, using longitudinal repeated measures of the frailty index: A prospective cohort study. *Age Ageing* **49**(6), 966–973 (2020).
34. Xu, W., Liang, Y. & Lin, Z. Association between Neutrophil-Lymphocyte ratio and Frailty: The Chinese longitudinal healthy longevity survey. *Front. Med. (Lausanne)* **8**, 783077 (2021).
35. Zhang, J., Wang, Q., Hao, W. & Zhu, D. Long-term food variety and dietary patterns are associated with frailty among Chinese older adults: A cohort study based on CLHLS from 2014 to 2018. *Nutrients* **14**(20) (2022).
36. Slaets, J. P. Vulnerability in the elderly: Frailty. *Med. Clin. North. Am.* **90**(4), 593–601 (2006).
37. Rockwood, K., Fox, R. A., Stolee, P., Robertson, D. & Beattie, B. L. Frailty in elderly people: An evolving concept. *Cmaj* **150**(4), 489–495 (1994).
38. Rockwood, K., Mogilner, A. & Mitnitski, A. Changes with age in the distribution of a frailty index. *Mech. Ageing Dev.* **125**(7), 517–519 (2004).

39. Mitnitski, A. B., Graham, J. E., Mogilner, A. J. & Rockwood, K. Frailty, fitness and late-life mortality in relation to chronological and biological age. *BMC Geriatr.* **2**, 1 (2002).
40. Kulminski, A. et al. Cumulative index of health disorders as an indicator of aging-associated processes in the elderly: Results from analyses of the National Long Term Care Survey. *Mech. Ageing Dev.* **128**(3), 250–258 (2007).
41. Woo, J., Goggins, W., Sham, A. & Ho, S. C. Public health significance of the frailty index. *Disabil. Rehabil.* **28**(8), 515–521 (2006).
42. American Medical Association white paper on elderly health. Report of the Council on Scientific affairs. *Arch. Intern. Med.* **150**(12), 2459–2472 (1990).

Acknowledgements

Author contributions
LH: Formal analysis, Writing - Original Draft. HC: Validation, Visualization, Data Curation. ZL: Conceptualization, Supervision, Project administration. All authors read and approved the final manuscript.

Funding

Declarations

Competing interests
The authors declare no competing interests.

Ethics approval and consent to participate
Not applicable since the dataset used in the study is publicly available. All methods were carried out in accordance with relevant guidelines and regulations.

Content for publication
Not applicable.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-74194-x.

Correspondence and requests for materials should be addressed to Z.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.