



# Explainable machine learning model for pre-frailty risk assessment in community-dwelling older adults

Chenlin Du<sup>1,2</sup>  | Zeyu Zhang<sup>2,3</sup> | Baoqin Liu<sup>3,4</sup> | Zijian Cao<sup>1,2</sup> |  
Nan Jiang<sup>1,2,3</sup>  | Zongjiu Zhang<sup>1,2,3</sup>

<sup>1</sup>School of Biomedical Engineering, Tsinghua University, Beijing, China

<sup>2</sup>Tsinghua Medicine, Tsinghua University, Beijing, China

<sup>3</sup>Institute for Hospital Management, Tsinghua University, Beijing, China

<sup>4</sup>Department of Gynecology of Traditional Chinese Medicine, China-Japan Friendship Hospital, Beijing, China

## Correspondence

Nan Jiang and Zongjiu Zhang, Institute for Hospital Management, Tsinghua University, Haidian District, Beijing, 100084, China.  
Email: [jiangnan1221@tsinghua.edu.cn](mailto:jiangnan1221@tsinghua.edu.cn) and [zhangzongjiu@mail.tsinghua.edu.cn](mailto:zhangzongjiu@mail.tsinghua.edu.cn)

## Funding information

None

## Abstract

**Background:** Frailty in older adults is linked to increased risks and lower quality of life. Pre-frailty, a condition preceding frailty, is intervenable, but its determinants and assessment are challenging. This study aims to develop and validate an explainable machine learning model for pre-frailty risk assessment among community-dwelling older adults.

**Methods:** The study included 3141 adults aged 60 or above from the China Health and Retirement Longitudinal Study. Pre-frailty was characterized by one or two criteria from the physical frailty phenotype scale. We extracted 80 distinct features across seven dimensions to evaluate pre-frailty risk. A model was constructed using recursive feature elimination and a stacking-CatBoost distillation module on 80% of the sample and validated on a separate 20% holdout data set.

**Results:** The study used data from 2508 community-dwelling older adults (mean age, 67.24 years [range, 60–96]; 1215 [48.44%] females) to develop a pre-frailty risk assessment model. We selected 57 predictive features and built a distilled CatBoost model, which achieved the highest discrimination (AUROC: 0.7560 [95% CI: 0.7169, 0.7928]) on the 20% holdout data set. The living city, BMI, and peak expiratory flow (PEF) were the three most significant contributors to pre-frailty risk. Physical and environmental factors were the top 2 impactful feature dimensions.

**Conclusions:** An accurate and interpretable pre-frailty risk assessment framework using state-of-the-art machine learning techniques and explanation methods has been developed. Our framework incorporates a wide range of features and

**Abbreviations:** ANN, artificial neural network; AUROC, area under the receiver operator characteristic curve; AutoML, auto machine learning; CatBoost, Categorical Boosting; CHARLS, China Health and Retirement Longitudinal Study; CI, confidence interval; KD, knowledge distillation; LightGBM, Light Gradient-Boosting Machine; LR, logistics regression; ML, machine learning; PACIFIC, interPretable ACcurate and effiCient pre-Frallty Classification; PEF, peak expiratory flow; PFP, physical frailty phenotype; RF, Random Forest; RFE, recursive feature elimination; SAGE, shapley additive global explanation; SHAP, SHapley Additive exPlanations; XAI, explainable artificial intelligence; XGBoost, eXtreme Gradient Bboosting; XT, eXtreme randomized tree.

Chenlin Du, Zeyu Zhang, and Baoqin Liu contributed equally to this study.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2024 The Author(s). *Health Care Science* published by John Wiley & Sons Ltd on behalf of Tsinghua University Press.

determinants, allowing for a comprehensive and nuanced understanding of pre-frailty risk.

#### KEYWORDS

China Health and Retirement Longitudinal Study, Chinese community-dwelling older adults, explainable machine learning, pre-frailty, risk assessment

## 1 | BACKGROUND

Frailty is a complex age-related clinical condition characterized by a decline in physiological capacity across several organ systems, with a resultant increased susceptibility to stressors [1]. Older adults with frailty have an increased likelihood of unmet care needs, falls and fractures, hospitalizations, lower quality of life, and early mortality [2, 3]. Pre-frailty is an evident risk-state before the onset of clinically identifiable frailty [4]. In China, the rapid expansion of the aging population has led to an increasing prevalence of frailty and pre-frailty, with the prevalence of 10% and 43%, respectively [5]. Evidence suggests that comparing with frail individuals, the pre-frail older adults are more susceptible to intervention, which can prevent and delay frailty or even reverse the state from pre-frailty to robust [6]. Therefore, effective strategies that target early detection and prevention of pre-frailty in an aging population become China's national priority to reduce the condition burden at the level of both individual and health system.

The emergence of machine learning (ML) in the field of pre-frailty is a relatively recent development [7, 8]. It has long been recognized that pre-frailty in the clinical population is heterogeneous in terms of physical, psychosocial, demographic, and environmental characteristics. The profiling focuses on understanding the individual's unique combination of characteristics and their association with pre-frailty. Traditional pre-frailty assessment scales cannot determine or predict pre-frailty risk with an acceptable degree of precision or reliability. The development of ML in pre-frailty research represents an important methodological advance in the field of precision health care as it improves the accuracy and potential clinical utility [9–12]. Given the additional pressure of the COVID-19 pandemic on already overstretched healthcare services, potential gains in scalable quality of care and improvements to resource efficiency are appealing.

Despite their proven effect, real-world implementation of ML into frailty research still faces three barriers. First, although the research agenda of the determinants of frailty keeps expanding, most of the studies focus on the identification of frailty instead of more intervenable

pre-frail individuals. Second, we know little about the relative importance of individual characteristics, built environment, social-economic contexts, and social policies that have only been examined in a piecemeal fashion. It is important to understand which factors exert the most important influence on pre-frailty among a large number of candidate predictors. Third, due to a lack of algorithm transparency, limited interpretability, and theoretical frameworks, clinicians are reluctant to trust the tools being integrated into care settings. The clinical implications of ML and how these techniques might be applied for early detection and intervention have largely been overlooked.

In this study, we drew data from the China Health and Retirement Longitudinal Study (CHARLS) to identify the most important determinants of pre-frailty for Chinese older adults from a large set of factors at biological, individual, and community levels. We took advantage of the recursive feature selection method to estimate the importance of factors cutting across multiple dimensions. We focus on the practical aspects of how explainable artificial intelligence can be clinically useful to increase model interpretability.

## 2 | METHODS

### 2.1 | Data source

The data were collected from the baseline survey (2011–2012) of CHARLS [13], an ongoing longitudinal cohort study of nationally representative community-dwelling adults aged 45 or older from China with comprehensive biomedical, clinical, and sociodemographic information. All participants were provided with written consent and the study protocol was approved by the Ethical Review Committee at Peking University (IRB00001052–11015).

It had been previously reported that the missing data can impact the interpretation of the prediction model [14]. Therefore, to avoid the potential bias, the current study only included participants without missing information. The overall exclusion criteria were as follows: (1) younger than 60 years old, (2) did not complete 2 or more

components in physical frailty phenotype (PFP) scale [15], (3) categorized as frailty by PFP scale, and (4) missing information present among selected features. We randomly selected 20% of the included subjects as the test set while the rest of them (80%) as the training set.

## 2.2 | Frailty level measurements

The frailty level of each participant was measured by the PFP scale [16] (validated on CHARLS [15]), in which five elements included weakness, slowness, exhaustion, inactivity, and shrinking (Supporting Information S1: Tables S1 and S2). Individuals with no criteria met were categorized as robust; those with one or two were classified as prefrail and three to five were considered frail. Details of frailty measures and cutoffs are shown in the supplementary methods.

## 2.3 | Determinants of pre-frailty

Through extensive literature review and expert consultation (Z.Z. and J.N.), we summarized the following 80 features from seven dimensions [1] including laboratory measurements, physical factors, demographics, health behaviors, comorbidities and medical histories, health-care service utilization and environmental factors for the analysis. 80 features are detailed in the supplementary methods.

## 2.4 | Analytical plan

As demonstrated in Figure 1, we optimize the IMPACT framework [19] by introducing Categorical Boosting (CatBoost) [20] to present the interPretable ACcurate and effICient pre-FraIlty Classification (PACIFIC) framework, which enables analysis of categorical features. Moreover, with the help of knowledge distillation [21] (KD) we introduced a stacking-CatBoost distillation module (Figure 1b) to improve the model performance. In the module, we first train several different tree-based models including Random Forest (RF), eXtreme Gradient Boosting (XGBoost), Light Gradient-Boosting Machine (LightGBM), CatBoost, eXtreme randomized Tree (XT) with k-fold bagging [22] as base model and stacked them using ensemble selection [23]. Whereafter, the stacking model is distilled into a CatBoost model to enable efficient calculation, analyses of categorical features and feature interaction effects. Hyperparameters for different models were illustrated in Supporting Information S1: Table S3.

## 2.5 | Recursive feature elimination (RFE)

Several literatures [19, 24] have demonstrated the efficiency of task-aware supervised features selection. In the current study, to ensure the time efficiency, a LighGBM [25] with fast training and high accuracy is applied to recursively eliminate features. In detail, we generate a random feature which follows the standard normal distribution first and then train the model, obtaining the area under the receiver operator characteristic curve (AUROC). After then, we rank the feature importance with the help of permutation feature importance [26]. We then remove up to 5% of the total features whose feature importance is below the randomly generated feature from the bottom of the feature importance ranked list. A new model is then trained with the remaining features with AUROC evaluated. If the AUROC of the new model is not improved, we then randomly choose up to 5% features, with the sampling rate inversely proportional to their feature importance (feature with lowest feature importance has the highest probability being sampled). Once the AUROC of the model improved after removing the selected feature, the feature importance is ranked again using permutation feature importance with selected features removed. We end the RFE if the early stopping criteria have been met, or all the features set below the randomly generated feature have been evaluated without improving the model performance. We keep the k-fold bagging the same as the stacking model to ensure the data consistency [27].

## 2.6 | Classifier modeling

For the purposes of accurate classification and analysis of categorical feature [28], a CatBoost [20] distilled from a stacking model is implemented to model the pre-frailty risk. AutoGluon-Tabular (v 0.7.0), an auto machine learning (AutoML) framework, is utilized to ensure the accurate and efficient training of the stacking model from different tree-based individual model via stacking, bagging, boosting and weighted combination. Whereafter, we distill our stacking model (teacher model) into a CatBoost (student model) with soft targets [21]. To demonstrate the accuracy of the distilled CatBoost, we also train RF, XGBoost, LightGBM, origin CatBoost (trained directly without knowledge distillation), XT, artificial neural network (ANN), and logistics regression (LR) for comparison. The hyperparameters (Supporting Information S1: Table S3) for each model are chosen by Random Search with 100 iterations and fivefold cross-validation. Model performance is measured with the AUROC, with accuracy, F1 score, precision and recall reported. The results

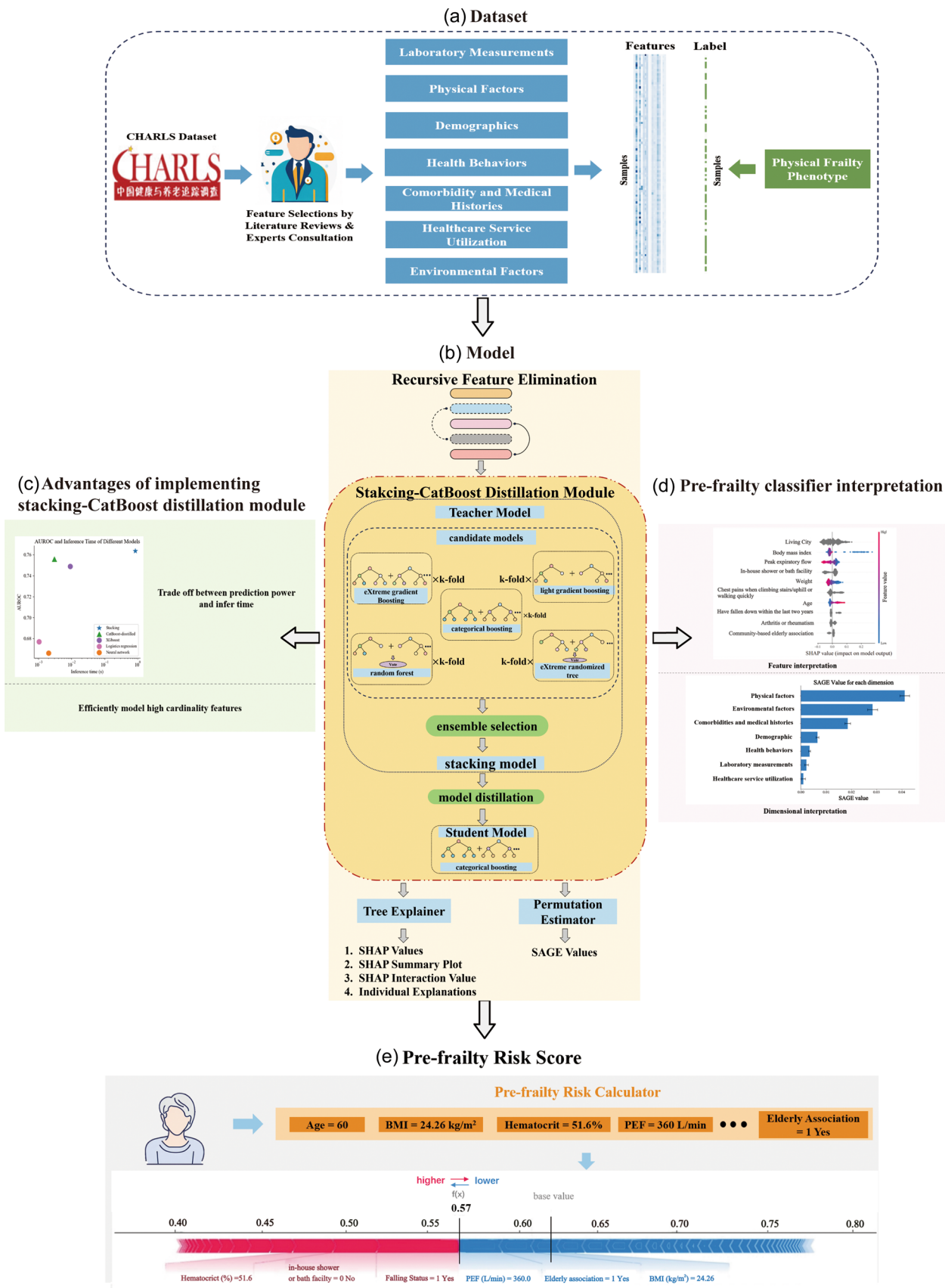


FIGURE 1 (See caption on next page).



are reported with mean values and a 95% confidence interval (CI) obtained from 1000 bootstrap samples for the test set. All individual models are built using the Scikit-learn package in Python 3.10. Simultaneously, we train the stacking and distilled CatBoost using AutoGluon-Tabular package. Note that, we keep the hyperparameters of distilled CatBoost the same as the directly trained CatBoost to evaluate the performance of KD. We also compare the infer time of stacking model, distilled CatBoost, XGBoost, ANN and LR to demonstrate the trade-off between model performance and computational efficiency.

## 2.7 | Performance evaluation

To address the robust of the RFE methods, accuracy, AUROC, F1 score, precision and recall rate of the stacking model on the test set before and after the feature selection was reported. We also compare the accuracy, AUROC, F1 score, precision, and recall rate of stacking model, distilled CatBoost (trained with stacking-CatBoost distillation module), RF, XGBoost, LightGBM, origin CatBoost (trained directly without model distillation), ANN, XT and LR. The hyperparameters of the above models were chosen by RandomSearch and fivefold cross-validation.

## 3 | RESULTS

### 3.1 | Participant characteristics

During the baseline survey of CHARLS, 17,708 Chinese residents participated, of whom 7681 were equal or older than 60 years old, and 5357 participants had data on at least four components. Participants classified as frailty ( $n = 407$ ) or containing missing data ( $n = 1819$ ) were also excluded from the study. Eventually, 3141 residents were included in the study, and 1926 of them were categorized as prefrailty and 1215 were identified as robustness.

### 3.2 | Feature selection

After 318 rounds of RFE (Supporting Information S1: Figure S1), the total number of features utilized in our stacking model dropped from 80 features to 57 features (Supporting Information S1: Table S4). Our results also demonstrated that the task aware RFE methods can significantly decrease the number of features while simultaneously improving the model performance of the stacking model (AUROC: 0.7599 [95% CI: 0.7222, 0.7965] vs. 0.7638 [95% CI: 0.7251, 0.8004]). We also evaluated the performance of individual models including RF, XGBoost, LighGBM, CatBoost, XT, ANN, LR before (Table 1) and after (Table 2) the feature selection and only LR reports a worsen model performance after feature pruning.

### 3.3 | Model performance

In the current study, stacking model outperforms all individual model before (AUROC: 0.7599 [95% CI: 0.7222, 0.7965]) and after (AUROC: 0.7638 [95% CI: 0.7251, 0.8004]) the RFE. Our result demonstrates that despite the highest AUROC is obtained by Catboost among all individual models, the stacking-CatBoost distillation module can still further improve the model performance with a statistic difference ( $p = 0.005$ ). Eventually, the CatBoost model reaches a similar model performance compared with stacking model, while significantly reducing the inference time (Supporting Information S1: Figure S2).

### 3.4 | Model explanation

Figure 2a presents a SHAP summary plot, which illustrates the magnitude, prevalence, and direction of the top 10 features with the greatest impact on the distilled CatBoost model. The living city, BMI, and peak expiratory flow (PEF) were identified as the three most significant contributors to the risk of pre-frailty.

---

**FIGURE 1** Overview of the PACIFIC framework. (a) We use the CHARLS (2011) data set, which contains 80 features that can be grouped into seven dimensions: laboratory measurements, physical factors, demographic, health behaviors, comorbidities and medical history, health-care services utilization, and environmental factors. (b) The PACIFIC framework combines an efficient recursive feature elimination method with a stacking-CatBoost distillation module and explainable AI techniques. (c) The advantages of the stacking-CatBoost distillation model. (d) The model interpretation of the PACIFIC framework. (e) The individualized explanations of PACIFIC show the impact of each risk factor on the overall pre-frailty risk score. In our framework, we implemented the Tree Explainer [17] to calculate SHapley Additive exPlanations values, providing local explanations for the distilled CatBoost model and aiding clinicians in analyzing its credibility. We also quantified the dimensional contribution of selected features to the model performance using Shapley Additive Global Explanation values, estimated by the Permutation Estimator [18].

**TABLE 1** Performance comparison of 80-feature model between RF, XGBoost, LightGBM, directly trained CatBoost, XT, ANN, LR, stacking model and distilled CatBoost.

	RF	XGBoost	LightGBM	CatBoost	XT	ANN	LR	Stacking	Distilled CatBoost
Accuracy	0.6864 [0.6491, 0.7241]	0.6805 [0.6443, 0.7177]	0.6732 [0.6363, 0.7113]	0.6965 [0.6587, 0.7321]	0.6599 [0.6220, 0.6970]	0.6209 [0.5837, 0.6603]	0.6574 [0.6204, 0.6938]	0.7159 [0.6794, 0.7528]	<b>0.6968</b> [0.6603, 0.7337]
AUROC	0.7292 [0.6882, 0.7669]	0.7380 [0.6979, 0.7764]	0.7127 [0.6690, 0.7526]	0.7492 [0.7095, 0.7857]	0.7103 [0.6693, 0.7508]	0.6347 [0.5906, 0.6787]	0.6792 [0.6353, 0.7191]	0.7599 [0.7222, 0.7965]	<b>0.7532</b> [0.7142, 0.7886]
F1 Score	0.7831 [0.7522, 0.8116]	0.7540 [0.7193, 0.7869]	0.7777 [0.7467, 0.8079]	0.7646 [0.7309, 0.7972]	0.7715 [0.7414, 0.8021]	0.7028 [0.6633, 0.7399]	0.7507 [0.7168, 0.7814]	<b>0.7865</b> [0.7531, 0.8167]	0.7733 [0.7412, 0.8033]
Precision	0.6956 [0.6544, 0.7343]	0.7342 [0.6905, 0.7786]	0.6821 [0.6425, 0.7216]	0.7499 [0.7087, 0.7928]	0.6703 [0.6292, 0.7122]	0.6965 [0.6517, 0.7438]	0.6949 [0.6509, 0.7382]	<b>0.7488</b> [0.7075, 0.7895]	0.7331 [0.6902, 0.7729]
Recall	<b>0.8964</b> [0.8626, 0.926]	0.7753 [0.7343, 0.8158]	0.9051 [0.8735, 0.9355]	0.7804 [0.7393, 0.8198]	0.9093 [0.8782, 0.935]	0.7097 [0.6650, 0.7532]	0.8167 [0.7763, 0.8561]	0.8288 [0.7906, 0.8656]	0.8186 [0.7815, 0.8553]

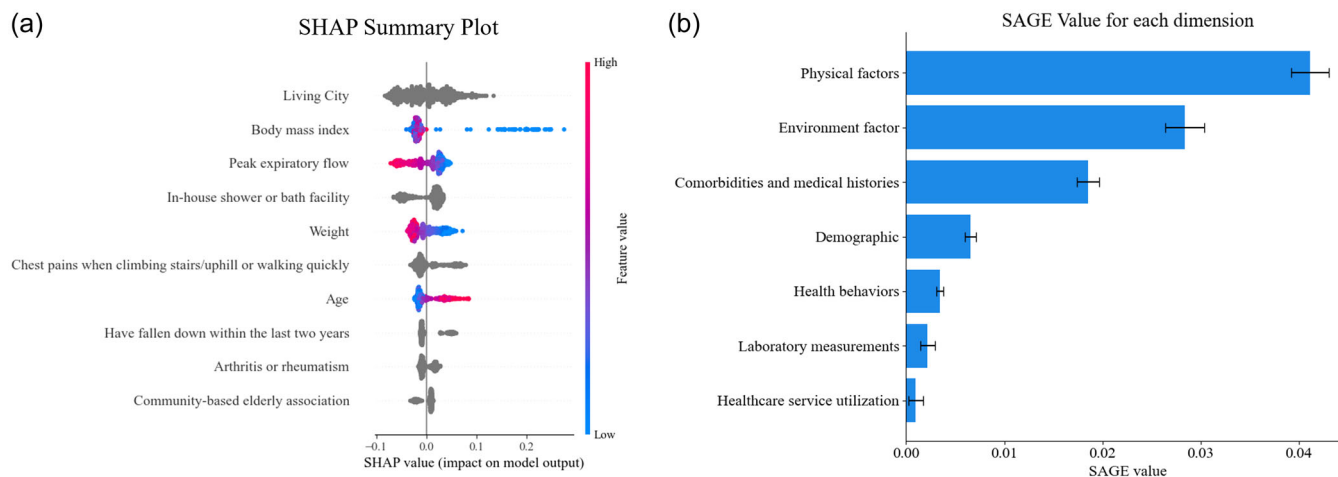
Abbreviation: AUROC, area under the receiver operator characteristic curve.

Note: The *p*-values for all tests are less than 0.001, indicating the significance of the values highlighted in bold.

**TABLE 2** Performance comparison of 57-feature model between RF, XGBoost, LightGBM, directly trained CatBoost, XT, ANN, LR, stacking model and distilled CatBoost.

	RF	XGBoost	LightGBM	CatBoost	XT	ANN	LR	Stacking	Distilled CatBoost
Accuracy	0.6915 [0.6539, 0.7273]	0.6853 [0.6475, 0.7193]	0.6653 [0.6268, 0.7018]	0.7125 [0.6762, 0.7480]	0.6694 [0.6316, 0.7065]	0.6395 [0.6028, 0.6762]	0.6446 [0.6061, 0.6826]	0.7203 [0.6842, 0.7576]	<b>0.7174</b> [0.6810, 0.7528]
AUROC	0.7448 [0.7057, 0.7828]	0.7489 [0.7099, 0.7855]	0.7152 [0.6738, 0.7536]	0.7537 [0.7152, 0.7911]	0.7225 [0.6820, 0.7621]	0.6662 [0.6238, 0.7083]	0.6770 [0.6328, 0.7173]	0.7638 [0.7251, 0.8004]	<b>0.7560</b> [0.7169, 0.7928]
F1 Score	0.7860 [0.7539, 0.8156]	0.7637 [0.7310, 0.7948]	0.7358 [0.6992, 0.7692]	0.7844 [0.7525, 0.8144]	0.7765 [0.7448, 0.8054]	0.7168 [0.6803, 0.7518]	0.7414 [0.7077, 0.7738]	0.7906 [0.7581, 0.8217]	<b>0.7884</b> [0.7555, 0.8172]
Precision	0.6998 [0.6596, 0.7402]	0.7266 [0.6839, 0.7689]	0.7341 [0.6889, 0.7767]	0.7455 [0.7037, 0.7878]	0.6779 [0.6370, 0.7181]	0.7117 [0.6658, 0.7551]	0.6863 [0.6423, 0.7288]	0.7503 [0.7061, 0.7922]	<b>0.7483</b> [0.7057, 0.7888]
Recall	<b>0.8967</b> [0.8646, 0.9239]	0.8054 [0.7674, 0.8444]	0.7381 [0.6925, 0.7789]	0.8282 [0.7876, 0.8653]	0.9092 [0.8772, 0.9361]	0.7225 [0.6780, 0.7685]	0.8067 [0.7649, 0.8435]	0.8359 [0.7980, 0.8727]	0.8334 [0.7961, 0.8663]

Note: The *p*-values for all tests are less than 0.001, indicating the significance of the values highlighted in bold.



**FIGURE 2** Global Risk Classification Using the PACIFIC Framework. (a) SHAP summary plot of the PACIFIC framework trained on the pre-frailty classification task. The plot displays the 10 most impactful features on classification, ranked from high to low. Numeric features are colored with red representing a higher feature value and blue indicating a lower value, while categorical or bool features are labeled in gray. A negative SHAP value, extending to the left, reduces the pre-frailty risk, while a positive value indicates an increased risk. (b) SAGE value plot for each feature dimension. Our results demonstrate that living city is the most influential feature in contributing to the pre-frailty risk through geographic variations, while physical factors are identified as the most impactful dimension. SAGE, Shapley Additive Global Explanation; SHAP, SHapley Additive exPlanations.

SHAP values also indicate the associations between features and the pre-frailty risk (Supporting Information S1: Figure S4). Living in certain cities and lacking community-based older adults associations increase pre-frailty risk. Conversely, living in a house with a household-installed water heater, and being younger with higher BMI, PEF, and weight decrease pre-frailty risk. Chest pains during physical exertion fall within the last 2 years, and arthritis or rheumatism diagnoses are the top three comorbidities that increase pre-frailty risk.

Figure 2b illustrates the contribution of various dimensions to the model performance of the distilled CatBoost. Physical factors, including BMI, PEF, weight, height, DBP, SBP, and pulse, are identified as the most impactful dimension, while healthcare service utilization, such as outpatient and hospitalization reasons, is found to be the least impactful dimension.

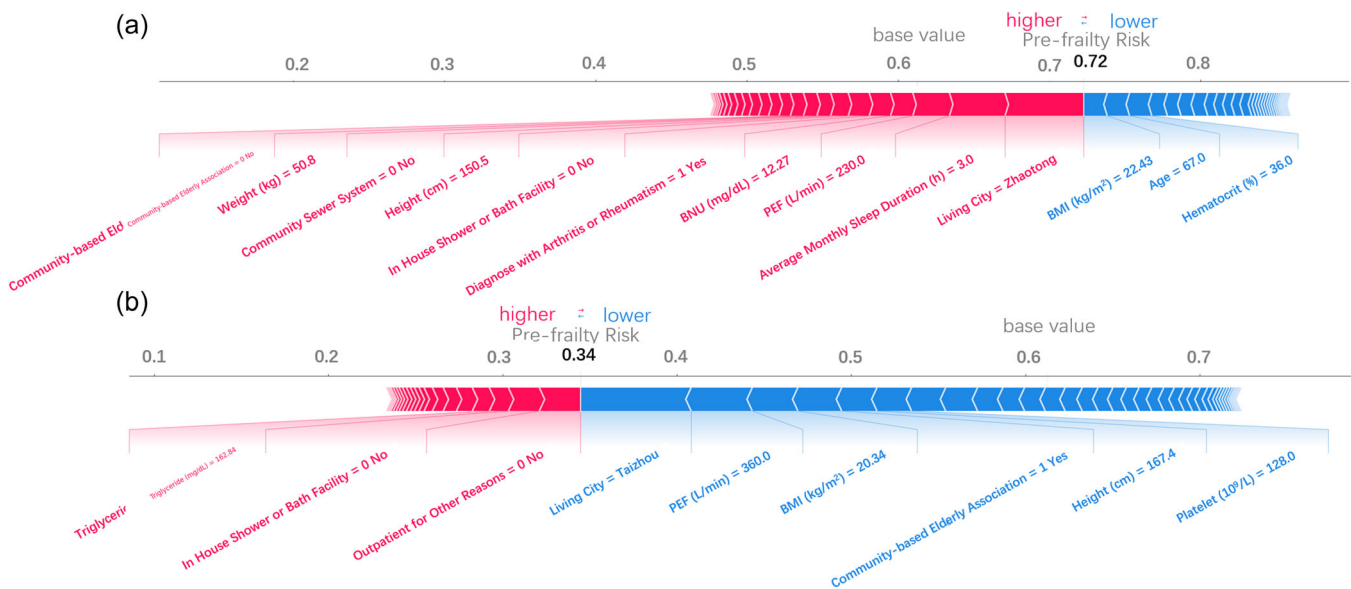
### 3.5 | Interpretable individual pre-frailty scores

As shown in Figure 3, the PACIFIC framework can visualize personal risk using features with the help of TreeExplainer and force plot. Figure 3a shows a pre-frail individual with a pre-frailty risk score of 0.72. Factors such as living in Zhaotong, a relatively underdeveloped mountainous city in southwestern China, without community-based older adult associations, community sewer systems, or in-house shower/bath facilities, as well

as short average monthly sleep duration (3 h), relatively low PEF (230 L/min), height (150.5 cm), weight (50.8 kg), and BUN (12.27 mg/dL) significantly increased the risk despite the individual's relatively young age (67), ideal BMI (22.43 kg/m<sup>2</sup>), and hematocrit (36%). Figure 3b shows a robust individual with a pre-frailty risk score of 0.34. Factors such as living in Taizhou, a relatively developed coastal city in southeastern China, with community-based older adult associations and ideal PEF (360 L/min), BMI (20.34 kg/m<sup>2</sup>), height (167.4 cm), and platelet count ( $128 \times 10^9/L$ ) decreased the risk, while outpatient treatment, relatively high triglyceride levels (162.84 mg/dL), and the absence of in-house shower/bath facilities increased the risk.

## 4 | DISCUSSION

In our study, we established an explainable machine learning framework (PACIFIC) to identify pre-frail older adults living in the community using the information on laboratory measurements, physical factors, demographics, health behaviors, comorbidities and medical histories, healthcare service utilization, household facilities, and neighborhood resources. This represents the first attempt to integrate state-of-the-art machine learning techniques and explanation methods to systematically study pre-frailty. We have enhanced the practical value of our study by introducing an easy-to-use and explainable pre-frailty risk calculator.



**FIGURE 3** Individual pre-frailty risk score calculator. The output value is the risk score for the individual. The base value is the mean risk score, which is the output without any features inputted. Features in red increase the risk, while features in blue decrease the risk. (a) A pre-frail individual with a risk score of 0.72 calculated by the pre-frailty risk calculator. (b) A robust individual with a risk score of 0.34 calculated by the pre-frailty risk calculator. The risk score calculator can also be used to assess the impact of improving specific indicators on the overall risk.

Our study demonstrates that distilled Catboost is an effective ML model for identifying pre-frailty from community-dwelling older adults. In our study, distilled Catboost surpassed all other individual models. Only a few studies [7, 8] have explored the feasibility of identifying pre-frailty from community-dwelling old adults using ML. We significantly improved the model performance (AUROC: 0.7560 vs. 0.7220 [7]) using the PACIFIC framework by leveraging the advantages of CatBoost and KD. Gradient boosting decision tree algorithms are known to efficiently model the heterogeneous data [28]. Among these models, CatBoost outperforms others [20, 28], due to its built-in ordered target statistic [20] that effectively processes categorical features common in modern data-driven medical problems [29]. We also utilized KD to boost our model performance. Although stacking models [23] can significantly reduce variance and improve performances, it increases inference time and is difficult to deploy. KD [21] can bridge the knowledge gap between stacking (teacher) and individual (student) models (Supporting Information S1: Figure S3), trading off between model performance and inference time (Supporting Information S1: Figure S2).

Machine learning models such as distilled CatBoost provide superior risk estimates, but the source of risk is often non-interpretable [17]. To overcome this issue, we visualized global and individual risk sources using SHAP summary plot (Figure 2a) and SHAP dependence plot

(Supporting Information S1: Figure S4), making our study one of the most comprehensive clinical applications with SHAP values. We identified the top 10 most impactful features and their associations with pre-frailty risk, most of which have been reported by prior epidemiological studies, including BMI [30, 31], PEF [32, 33], weight [34], age [35], falls within the last 2 years [36], and arthritis or rheumatism [37, 38] (Supporting Information S1: Figure S3). However, their complex interactions result in varying risk contributions. With the aid of a SHAP individual explanation plot (Figure 3), we can visually discern the risk contributions of each feature, helping individuals improve their self-awareness [19] or providing potential avenues for risk modification [39].

We employed feature selection to effectively eliminate less predictive variables and methodically examine potential pre-frailty risk factors from various sources. This led to the identification of previously overlooked risk factors, including chest pains during physical exertion, the presence of in-house shower or bath facilities, and participation in community-based older adult associations. Chest pains during physical exertion have been recognized as a significant indicator of acute coronary syndrome [40], which is closely linked to pre-frailty [41]. However, the connection between chest discomfort during physical activity and frailty remains unclear. Meanwhile, few studies have explored the determinants of household amenities and neighborhood resources [42]. Our research revealed that the absence of in-home shower or bath amenities or



community-based senior citizen organizations increased the risk of pre-frailty. The absence of in-home shower or bath amenities may hinder personal hygiene or reflect substandard indoor conditions, increasing the risk of pre-frailty. On the other hand, senior citizen organizations offer opportunities for social interaction [43], physical exercise [44], and access to health information and resources [45], aiding in the prevention of pre-frailty. From a global standpoint, our study highlighted the significance of environmental factors using SAGE values (Figure 2b), which provide comparable contributions that aid in model explanation and knowledge discovery in clinical practice [46, 47]. Our findings align with recent arguments that environmental factors play a crucial role in health longevity [45] and frailty [42, 48].

One major drawback of traditional epidemiological research is the difficulty of modeling high cardinality features, which are frequently one-hot encoded (dummy) into sparse vectors [25], leading to an underestimate of their importance. However, the build-in algorithm of CatBoost efficiently handles these features [20] and, in the PACIFIC framework, ranked living city as the most influential feature (Figure 2a) in contributing to the pre-frailty risk through geographic variations. This finding aligns with previous studies [15, 49] that reported regional disparities in healthcare as a cause for geographic variation in the prevalence of frailty among Chinese community-dwelling older adults. This could be attributed to several underlying factors, such as geographic factors (i.e., altitudes and climates) [50], imbalanced healthcare facilities, and service of different cities [51], as well as varying average city education and income level [52]. Our findings underscore the importance of considering geographical variations in public health strategies and interventions aimed at reducing pre-frailty among older adults.

To our knowledge, this was also the first study using a machine learning-based method to summarize the dimensional risk based on sources of risk. The SAGE values enabled us to visualize the global impact of feature dimension and quantify their importance. This is critical not only for clinical counseling but also for the development of optimized preventative strategies that are personalized and standardized. The early identification of pre-frailty can significantly reduce the onset of frailty and reduce healthcare costs. There is emerging evidence that indicates early prevention needs more person-centered community care planning [53]. It is also known that these older adults represent a heterogeneous cohort, and a one-size-fits-all approach is unlikely to be successful. For practice, a comprehensive assessment could be distilled to a 4-layer system: environmental factors, social factors, clinical factors, and biological factors. This

approach combines system-level targets with the development of person-centered healthcare, which has been little studied. Our model may provide earlier alerts compared to commonly used assessment tools, such as Activities of Daily Living evaluations and serve as a valuable self-assessment or clinical decision support tool. The integration of our model with existing frameworks could enhance its clinical utility and facilitate broader adoption. Future work is needed to validate the effectiveness of our frameworks.

The best timing for using ML as a clinical decision-support tool might be embedded within the physical examination and outpatient services, for example, where an on-screen notification during a consultation informs the clinician about their patient's pre-frailty risk. These notifications might be particularly beneficial in healthcare settings in which a patient does not always see the same clinician, and the clinician might not be aware of a patient's full history and living environment.

Our research has several limitations. First, although BMI is an important criterion in the PFP, we did not exclude it due to its cost-efficiency and our aim to analyze the effect of high BMI on the risk of pre-frailty. Second, due to the lack of information on shrinking [54] and neighborhood information [42], we only analyzed the baseline data, which introduces potential bias and limits our study to a cross-sectional analysis, despite our framework's potential for extension by modifying the outcome. Third, although we have rigorously validated the model performance on a holdout unseen set, external validations are still required to ensure the robustness of our model. Fourth, we included 80 features from seven dimensions based on existing literature and experts' consultation, the list is not exhaustive. Meanwhile, the relationship detected by our study cannot be claimed to be causal, and further studies are required. Nonetheless, the study demonstrated the utility of the state-of-the-art ML and explanation method in accurate and interpretable pre-frailty risk quantification which may facilitate the risk stratification and residency consultation. The identification and understanding of several underappreciated determinants in our framework may also have important implications in policymaking and intervention design.

## 5 | CONCLUSIONS

In conclusion, we have developed an accurate and interpretable pre-frailty risk assessment framework using state-of-the-art ML techniques and explanation methods. Our framework incorporates a wide range of features and determinants, allowing for a comprehensive and nuanced understanding of pre-frailty risk. The results of

our study demonstrate the potential for ML-based approaches to improve risk stratification and inform targeted interventions in the field of epidemiology. Beyond the current study, the PACIFIC framework holds promise for adaptation to other populations and health conditions, thereby contributing to the formulation of broader public health strategies. Further research is needed to validate our findings and explore the potential for wider application of our framework.

## AUTHOR CONTRIBUTIONS

**Chenlin Du:** Conceptualization (equal); data curation (equal); software (equal), validation (equal); visualization (equal); writing—original draft (equal). **Zeyu Zhang:** Conceptualization (equal). **Baoqin Liu:** Investigation (equal). **Zijian Cao:** Investigation (equal). **Nan Jiang:** Writing—original draft (equal); writing—review & editing (equal). **Zongjiu Zhang:** Funding acquisition (equal); project administration (equal); supervision (equal).

## ACKNOWLEDGMENTS

None.

## CONFLICT OF INTEREST STATEMENT

Professor Zongjiu Zhang is a member of the *Health Care Science* Editorial Board. To minimize bias, he was excluded from all editorial decision-making related to the acceptance of this article for publication. The remaining authors declare no conflict of interest.

## DATA AVAILABILITY STATEMENT

No new data are generated in this study. The CHARLS data is available at <http://charls.pku.edu.cn/>.

## ETHICS STATEMENT

The study protocol was approved by the Ethical Review Committee at Peking University (IRB00001052–11015).

## INFORMED CONSENT

All participants were provided with written consent.

## ORCID

Chenlin Du  <http://orcid.org/0000-0001-7039-8542>

Nan Jiang  <http://orcid.org/0000-0003-1469-0375>

## REFERENCES

1. Hoogendijk EO, Afilalo J, Ensrud KE, Kowal P, Onder G, Fried LP. Frailty: implications for clinical practice and public health. *Lancet*. 2019;394(10206):1365–75. [https://doi.org/10.1016/S0140-6736\(19\)31786-6](https://doi.org/10.1016/S0140-6736(19)31786-6)
2. Fan J, Yu C, Guo Y, Bian Z, Sun Z, Yang L, et al. Frailty index and all-cause and cause-specific mortality in Chinese adults: a prospective cohort study. *Lancet Public Health*. 2020;5(12):e650–60. [https://doi.org/10.1016/S2468-2667\(20\)30113-4](https://doi.org/10.1016/S2468-2667(20)30113-4)
3. Fan L, Tian Y, Wang J, Ding Y, Wang S, Xue H, et al. Frailty predicts increased health care utilization among community-dwelling older adults: a longitudinal study in China. *J Am Med Dir Assoc*. 2021;22(9):1819–24. <https://doi.org/10.1016/j.jamda.2021.01.082>
4. Sezgin D, Liew A, O'Donovan MR, O'Caomh R. Pre-frailty as a multi-dimensional construct: a systematic review of definitions in the scientific literature. *Geriatr Nurs (Minneapolis)*. 2020;41(2):139–46. <https://doi.org/10.1016/j.gerinurse.2019.08.004>
5. He B, Ma Y, Wang C, Jiang M, Geng C, Chang X, et al. Prevalence and risk factors for frailty among community-dwelling older people in China: a systematic review and meta-analysis. *J Nutr Health Aging*. 2019;23(5):442–50. <https://doi.org/10.1007/s12603-019-1179-9>
6. Gené Huguet L, Navarro González M, Kostov B, Ortega Carmona M, Colungo Francia C, Carpallo Nieto M, et al. Pre frail 80: multifactorial intervention to prevent progression of pre-frailty to frailty in the elderly. *J Nutr Health Aging*. 2018;22(10):1266–74. <https://doi.org/10.1007/s12603-018-1089-2>
7. Sajeev S, Champion S, Maeder A, Gordon S. Machine learning models for identifying pre-frailty in community dwelling older adults. *BMC Geriatr*. 2022;22(1):794. <https://doi.org/10.1186/s12877-022-03475-9>
8. Teh SK, Rawtaer I, Tan HP. Predictive accuracy of digital biomarker technologies for detection of mild cognitive impairment and pre-frailty amongst older adults: a systematic review and meta-analysis. *IEEE J Biomed Health Inform*. 2022;26(8):3638–48. <https://doi.org/10.1109/JBHI.2022.3185798>
9. Akbari G, Nikkhoo M, Wang L, Chen CPC, Han DS, Lin YH, et al. Frailty level classification of the community elderly using microsoft kinect-based skeleton pose: a machine learning approach. *Sensors*. 2021;21(12):4017. <https://doi.org/10.3390/s21124017>
10. Ju C, Zhou J, Lee S, Tan MS, Liu T, Bazoukis G, et al. Derivation of an electronic frailty index for predicting short-term mortality in heart failure: a machine learning approach. *ESC Heart Failure*. 2021;8(4):2837–45. <https://doi.org/10.1002/ehf2.13358>
11. Lv W, Liao H, Wang X, Yu S, Peng Y, Li X, et al. A machine learning-based assistant tool for early frailty screening of patients receiving maintenance hemodialysis. *Int Urol Nephrol*. 2024;56(1):223–35. <https://doi.org/10.1007/s11255-023-03640-y>
12. Wu Y, Jia M, Xiang C, Fang Y. Latent trajectories of frailty and risk prediction models among geriatric community dwellers: an interpretable machine learning perspective. *BMC Geriatr*. 2022;22(1):900. <https://doi.org/10.1186/s12877-022-03576-5>
13. 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. <https://doi.org/10.1093/ije/dys203>
14. Payrovnaziri SN, Xing A, Salman S, Liu X, Bian J, He Z. Assessing the impact of imputation on the interpretations of prediction models: a case study on mortality prediction for patients with acute myocardial infarction. *AMIA Joint Summits Transl Sci*. 2021;2021:465–74.
15. Wu C, Smit E, Xue QL, Odden MC. Prevalence and correlates of frailty among community-dwelling Chinese older adults: the China health and retirement longitudinal study. *J Gerontol: Ser A*. 2018;73(1):102–8. <https://doi.org/10.1093/gerona/glx098>

16. Fried LP, Tangen CM, Walston J, Newman AB, Hirsch C, Gottdiener J, et al. Frailty in older adults: evidence for a phenotype. *J Gerontol A Biol Sci Med Sci*. 2001;56(3):M146–57. <https://doi.org/10.1093/gerona/56.3.m146>
17. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From local explanations to global understanding with explainable AI for trees. *Nat Mach Intell*. 2020;2(1):56–67. <https://doi.org/10.1038/s42256-019-0138-9>
18. Covert I, Lundberg S, Lee SI. Explaining by removing: a unified framework for model explanation. 2020. <http://arxiv.org/abs/2011.14878>
19. Qiu W, Chen H, Dincer AB, Lundberg S, Kaeberlein M, Lee SI. Interpretable machine learning prediction of all-cause mortality. *Commun Med*. 2022;2:125. <https://doi.org/10.1038/s43856-022-00180-x>
20. Prokhorenkova L, Gusev G, Vorobev A., Dorogush A. V., Gulin A. CatBoost: unbiased boosting with categorical features. In: *Proceedings of the 32nd International Conference on Neural Information Processing Systems*. Montréal, Canada: Curran Associates Inc; 2018. p. 6639–49. <https://doi.org/10.5555/3327757.3327770>
21. Fakoor R, Mueller J, Erickson N, Chaudhari P, Smola AJ. Fast, accurate, and simple models for tabular data via augmented distillation. In: *Proceedings of the 34th International Conference on Neural Information Processing Systems*. Vancouver, BC, Canada: Curran Associates Inc; 2020. p. 727. <https://doi.org/10.5555/3495724.3496451>
22. Parmanto B, Munro PW, Doyle HR. Reducing variance of committee prediction with resampling techniques. *Connection Sci*. 1996;8(3–4):405–26. <https://doi.org/10.1080/095400996116848>
23. Caruana R, Niculescu-Mizil A, Crew G, Ksikes A. Ensemble selection from libraries of models. In: *Proceedings of the twenty-first international conference on Machine learning*. Banff, Alberta, Canada: Association for Computing Machinery; 2004. p. 18. <https://doi.org/10.1145/1015330.1015432>
24. Roy D, Murty KSR, Mohan CK. Feature selection using Deep Neural Networks. 2015 International Joint Conference on Neural Networks (IJCNN), Killarney, Ireland, 2015:1–6. <https://doi.org/10.1109/IJCNN.2015.7280626>
25. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al. LightGBM: a highly efficient gradient boosting decision tree. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Long Beach, California, USA: Curran Associates Inc; 2017. p. 3149–57. <https://doi.org/10.5555/3294996.3295074>
26. Altmann A, Toloşi L, Sander O, Lengauer T. Permutation importance: a corrected feature importance measure. *Bioinformatics*. 2010;26(10):1340–7. <https://doi.org/10.1093/bioinformatics/btq134>
27. Erickson N, Mueller J, Shirkov A, Zhang H, Larroy P, Li M, et al. AutoGluon-tabular: robust and accurate AutoML for structured data. 2020. <http://arxiv.org/abs/2003.06505>
28. Hancock JT, Khoshgoftaar TM. CatBoost for big data: an interdisciplinary review. *J Big Data*. 2020;7(1):94. <https://doi.org/10.1186/s40537-020-00369-8>
29. Ramírez S, Quiroz AJ, Riascos AJ. A supervised clustering MCMC methodology for large categorical feature spaces. *Stat Methods Med Res*. 2021;30(7):1708–24. <https://doi.org/10.1177/09622802211009258>
30. Kim M, Lee Y, Kim EY, Park Y. Mediating effect of waist: height ratio on the association between BMI and frailty: the Korean frailty and aging cohort study. *Br J Nutr*. 2020;124(5):513–20. <https://doi.org/10.1017/s0007114519002058>
31. Lee Y, Kim J, Han ES, Ryu M, Cho Y, Chae S. Frailty and body mass index as predictors of 3-year mortality in older adults living in the community. *Gerontology*. 2014;60(6):475–82. <https://doi.org/10.1159/000362330>
32. Vaz Fragoso CA, Enright PL, McAvay G, Van Ness PH, Gill TM. Frailty and respiratory impairment in older persons. *Am J Med*. 2012;125(1):79–86. <https://doi.org/10.1016/j.amjmed.2011.06.024>
33. Magave JA, Bezerra SJS, Matos AP, Pinto ACPN, Pegorari MS, Ohara DG. Peak expiratory flow as an index of frailty syndrome in older adults: a cross-sectional study. *J Nutr Health Aging*. 2020;24(9):993–8. <https://doi.org/10.1007/s12603-020-1509-y>
34. Goisser S, Guyonnet S, Volkert D. The role of nutrition in frailty: an overview. *J Frailty Aging*. 2016;5(2):74–7. <https://doi.org/10.14283/jfa.2016.87>
35. Rockwood K. Changes with age in the distribution of a frailty index. *Mech Ageing Dev*. 2004;125(7):517–9. <https://doi.org/10.1016/j.mad.2004.05.003>
36. Liu LK, Lee WJ, Chen LY, Hwang AC, Lin MH, Peng LN, et al. Association between frailty, osteoporosis, falls and hip fractures among community-dwelling people aged 50 years and older in Taiwan: results from I-lan longitudinal aging study. *PLoS One*. 2015;10(9):e0136968. <https://doi.org/10.1371/journal.pone.0136968>
37. Salaffi F, Di Matteo A, Farah S, Di Carlo M. Inflammaging and frailty in immune-mediated rheumatic diseases: how to address and score the issue. *Clin Rev Allergy Immunol*. 2023;64(2):206–21. <https://doi.org/10.1007/s12016-022-08943-z>
38. Gao RC, Wu ZG, Wu ZZ, Hao M, Wu GC. Frailty in rheumatoid arthritis: a systematic review and meta-analysis. *Joint Bone Spine*. 2022;89(4):105343. <https://doi.org/10.1016/j.jbspin.2022.105343>
39. Deng H, Eftekhari Z, Carlin C, Veerapong J, Fournier KF, Johnston FM, et al. Development and validation of an explainable machine learning model for major complications after cytoreductive surgery. *JAMA Netw Open*. 2022;5(5):e2212930. <https://doi.org/10.1001/jamanetworkopen.2022.12930>
40. Swap CJ. Value and limitations of chest pain history in the evaluation of patients with suspected acute coronary syndromes. *JAMA*. 2005;294(20):2623–9. <https://doi.org/10.1001/jama.294.20.2623>
41. Sergi G, Veronese N, Fontana L, De Rui M, Bolzetta F, Zambon S, et al. Pre-frailty and risk of cardiovascular disease in elderly men and women. *J Am Coll Cardiol*. 2015;65(10):976–83. <https://doi.org/10.1016/j.jacc.2014.12.040>
42. Liu H, Chen B, Li Y, Morrow-Howell N. Neighborhood resources associated with frailty trajectories over time among community-dwelling older adults in China. *Health Place*. 2022;74:102738. <https://doi.org/10.1016/j.healthplace.2021.102738>
43. Ye B, Chen H, Huang L, Ruan Y, Qi S, Guo Y, et al. Changes in frailty among community-dwelling Chinese older adults and its predictors: evidence from a two-year longitudinal study. *BMC Geriatr*. 2020;20(1):130. <https://doi.org/10.1186/s12877-020-01530-x>
44. Trombetti A, Hars M, Hsu FC, Reid KF, Church TS, Gill TM, et al. Effect of physical activity on frailty: secondary analysis of

- a randomized controlled trial. *Ann Intern Med.* 2018;168(5):309–16. <https://doi.org/10.7326/M16-2011>
45. Chen X, Giles J, Yao Y, Yip W, Meng Q, Berkman L, et al. The path to healthy ageing in China: a Peking University-Lancet Commission. *Lancet.* 2022;400(10367):1967–2006. [https://doi.org/10.1016/S0140-6736\(22\)01546-X](https://doi.org/10.1016/S0140-6736(22)01546-X)
46. Erion G, Janizek JD, Hudelson C, Utarnachitt RB, McCoy AM, Sayre MR, et al. A cost-aware framework for the development of AI models for healthcare applications. *Nat Biomed Eng.* 2022;6(12):1384–98. <https://doi.org/10.1038/s41551-022-00872-8>
47. Janizek JD, Dincer AB, Celik S, Chen H, Chen W, Naxerova K, et al. Uncovering expression signatures of synergistic drug responses via ensembles of explainable machine-learning models. *Nat Biomed Eng.* 2023;7(6):811–29. <https://doi.org/10.1038/s41551-023-01034-0>
48. Zimmermann J, Hansen S, Wagner M. Home environment and frailty in very old adults. *Zeitschrift für Gerontologie und Geriatrie.* 2021;54(Suppl 2):114–9. <https://doi.org/10.1007/s00391-021-01969-6>
49. Ma L, Tang Z, Zhang L, Sun F, Li Y, Chan P. Prevalence of frailty and associated factors in the community-dwelling population of China. *J Am Geriatr Soc.* 2018;66(3):559–64. <https://doi.org/10.1111/jgs.15214>
50. Sinclair DR, Maharani A, Chandola T, Bower P, Hanratty B, Nazroo J, et al. Frailty among older adults and its distribution in England. *J Frailty Aging.* 2022;11(2):163–8. <https://doi.org/10.14283/jfa.2021.55>
51. China statistical yearbook, National bureau of statistics of the People's Republic of China. <https://www.stats.gov.cn/sj/ndsj/2011/indexeh.htm>
52. Qi X, Li Y, Hu J, Meng L, Zeng P, Shi J, et al. Prevalence of social frailty and its associated factors in the older Chinese population: a national cross-sectional study. *BMC Geriatr.* 2023;23(1):532. <https://doi.org/10.1186/s12877-023-04241-1>
53. Zacharaki EI, Deltouzos K, Kalogiannis S, Kalamaras I, Bianconi L, Degano C, et al. FrailSafe: an ICT platform for unobtrusive sensing of Multi-Domain frailty for personalized interventions. *IEEE J Biomed Health Inform.* 2020;24(6):1557–68. <https://doi.org/10.1109/JBHI.2020.2986918>
54. Sha S, Pan Y, Xu Y, Chen L. Associations between loneliness and frailty among older adults: evidence from the China Health and Retirement Longitudinal Study. *BMC Geriatr.* 2022;22(1):537. <https://doi.org/10.1186/s12877-022-03044-0>

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Du C, Zhang Z, Liu B, Cao Z, Jiang N, Zhang Z. Explainable machine learning model for pre-frailty risk assessment in community-dwelling older adults. *Health Care Sci.* 2024;3:426–37. <https://doi.org/10.1002/hcs2.120>