



OPEN Prediction of coronary heart disease based on klotho levels using machine learning

Yuan Yao^{1,5}, Ying Zhao^{1,5}, Haifeng Li¹, Yanlin Han¹, Yue Wu¹, Renwei Guo², Mingfeng Ma^{2,3}✉ & Lixia Bu⁴✉

The diagnostic accuracy for coronary heart disease (CHD) needs to be improved. Some studies have indicated that klotho protein levels upon admission comprise an independent risk factor for CHD and have clinical value for predicting CHD. This study aimed to construct a tool to predict CHD risk by analyzing klotho levels and clinically relevant indicators by using a machine learning (ML) method. We randomly assigned the dataset of the National Health and Nutrition Examination Survey (NHANES) 2007–2016 to training and test sets at a ratio of 70:30. We evaluated the ability of five models constructed using logistic regression, neural networks, random forest, support vector machine, and eXtreme Gradient Boosting to predict CHD. We determined their predictive performance using the following parameters: area under the receiver operating characteristic curve, accuracy, precision, recall, F1, and Brier scores. We analyzed data from 11,583 persons in US NHANES and entered 13 potential predictive variables, including klotho and other clinically relevant indicators, into the feature screening process. We established that the five ML models could predict the onset of CHD. The RF model showed the best predictive performance among the five ML models.

Keywords Klotho, Machine learning, Coronary heart disease, Model

The recent increase in the incidence of coronary heart disease (CHD) has led to the critical state of global disease burden. Thus, reducing the global burden of CHD remains a public health priority¹. Studies have explored hypertension, hyperlipidemia, hyperglycemia, smoking, drinking, and obesity as risk factors for CHD². However, the association between klotho and CHD has not been investigated in detail^{3,4} probably because most studies have used traditional statistical analyses and because klotho biomarkers have not been widely applied in routine clinical practice. Therefore, a new analytical approach might be able to more accurately determine an association between CHD and klotho.

Klotho affects several metabolic pathways that are essential for the pathogenesis and prevention of cardiovascular diseases. It inhibits lipid peroxidation and inflammation and prevents endothelial injury and vascular calcification⁵. Some studies have found that klotho protein values are independent risk factors for CHD and have clinical value for predicting this disease⁶. However, the prediction of CHD onset using klotho has not been thoroughly investigated.

Machine learning (ML) broadly refers to the process of fitting predictive models to data or identifying informative groups within data. The aim of ML is to approximate or imitate the ability of humans to objectively recognize patterns using computation. ML is applied in almost all fields of biology⁷. We aimed to construct a prediction tool by analyzing klotho levels in patients with CHD, as well as values of clinically relevant indicators to predict the risk of CHD using ML.

Results

We assigned the 11,583 participants into groups with ($n = 1074$) and without ($n = 10,509$) CHD (Table 1). Figure 1 shows a flowchart of the participants through the study. Various CHD samples showed significant differences for all variables except alcohol consumption ($p < 0.05$). We entered 13 potential predictive variables into the feature screening process of the Least Absolute Shrinkage and Selection Operator regression. This statistical

¹Fenyang College of Shanxi Medical University, Fenyang 032200, Shanxi, China. ²Department of Cardiovascular Medicine, Fenyang Hospital Affiliated to Shanxi Medical University, Fenyang 032200, Shanxi, China. ³Department of Internal Medicine, Fenyang College of Shanxi Medical University, Fenyang 032200, Shanxi, China. ⁴Department of Geratology, Fenyang Hospital Affiliated to Shanxi Medical University, Fenyang 032200, Shanxi, China. ⁵Yuan Yao and Ying Zhao contributed equally to this work and share first authorship. ✉email: mamingfeng106@sina.com; Bulixia106@sina.com

Characteristics	Coronary_heart_disease (%)		Total	p
	Without CHD	With CHD		
Age years	57.0 (48.0, 65.0)	65.0 (58.8, 72.0)	57.7 (49.0, 66.0)	0.000**
Gender %				
Female	5512 (52.45)	391 (36.41)	5903 (50.96)	0.000**
Male	4997 (47.55)	683 (63.59)	5680 (49.04)	
Race %				
Mexican American	1645 (15.65)	122 (11.36)	1767 (15.26)	0.000**
Non-Hispanic Black	2089 (19.88)	198 (18.44)	2287 (19.74)	
Non-Hispanic White	4704 (44.76)	565 (52.61)	5269 (45.49)	
Other	2071 (19.71)	189 (17.60)	2260 (19.51)	
Education %				
Above high school	5484 (52.18)	450 (41.90)	5934 (51.23)	0.000**
High school or GED	2317 (22.05)	260 (24.21)	2577 (22.25)	
Less than high school	2708 (25.77)	364 (33.89)	3072 (26.52)	
Marital %				
Married/Living with partner	6835 (65.04)	650 (60.52)	7485 (64.62)	0.000**
Never married	879 (8.36)	68 (6.33)	947 (8.18)	
Widowed/Divorced/Separated	2795 (26.60)	356 (33.15)	3151 (27.20)	
PIR %				
High	3705 (35.26)	248 (23.09)	3953 (34.13)	0.000**
Low	3007 (28.61)	459 (42.74)	3466 (29.92)	
Middle	3797 (36.13)	367 (34.17)	4164 (35.95)	
BMI %				
Normal weight	2517 (23.95)	188 (17.50)	2705 (23.35)	0.000**
Obese	4316 (41.07)	557 (51.86)	4873 (42.07)	
Over weight	3676 (34.98)	329 (30.63)	4005 (34.58)	
Smoke %				
No	5478 (52.13)	367 (34.17)	5845 (50.46)	0.000**
Yes	5031 (47.87)	707 (65.83)	5738 (49.54)	
Alcohol %				
Current	7463 (71.02)	782 (72.81)	8245 (71.18)	0.073
Former	1537 (14.63)	165 (15.36)	1702 (14.69)	
Never	1509 (14.36)	127 (11.82)	1636 (14.12)	
DM %				
No	8230 (78.31)	591 (55.03)	8821 (76.15)	0.000**
Yes	2279 (21.69)	483 (44.97)	2762 (23.85)	
Hypertension %				
No	5146 (48.97)	228 (21.23)	5374 (46.40)	0.000**
Yes	5363 (51.03)	846 (78.77)	6209 (53.60)	
Hyperglycaemia %				
No	2326 (22.13)	138 (12.85)	2464 (21.27)	0.000**
Yes	8183 (77.87)	936 (87.15)	9119 (78.73)	
Klotho pg/ml	805.2 (657.4, 994.4)	773.1 (620.7, 946.4)	853.1 (654.5, 991.4)	0.000**

Table 1. Characteristics of the study participants with and without CHD from 2007–2016 in the US NHANES. CHD, coronary heart disease; DM, diabetes mellitus; NHANES, National Health and Nutrition Examination Survey; PIR, poverty to income ratio. * $p < 0.05$, ** $p < 0.01$.

approach uses dummy variable processing to process 13 variables into 20 variables. The initial lambda value was determined by 5-fold cross-validation, and 13 initial variables were selected (Figure S1). These comprised education, marital status, race, poverty impact ratio, sex, body mass index, smoking, alcohol consumption, diabetes mellitus, hypertension, hyperlipidemia, age, and klotho levels. These variables were randomly assigned to training and test sets at a ratio of 70:30, and the training set was resampled. After hyperparameter optimization, the logistic regression (LR), neural network (NNet), random forest (RF), support vector machine (SVM), and eXtreme Gradient Boosting (XGBoost) models were used to construct the model in the training set. LR is a traditional multivariate statistical learning model that predicts class output, which can be either “0” or “1.” LR is a popular ML algorithm, and it is grouped under supervised learning techniques⁸. Thus, we considered using LR to benchmark the predictive performance of the other ML models. The area under receiver operating characteristic

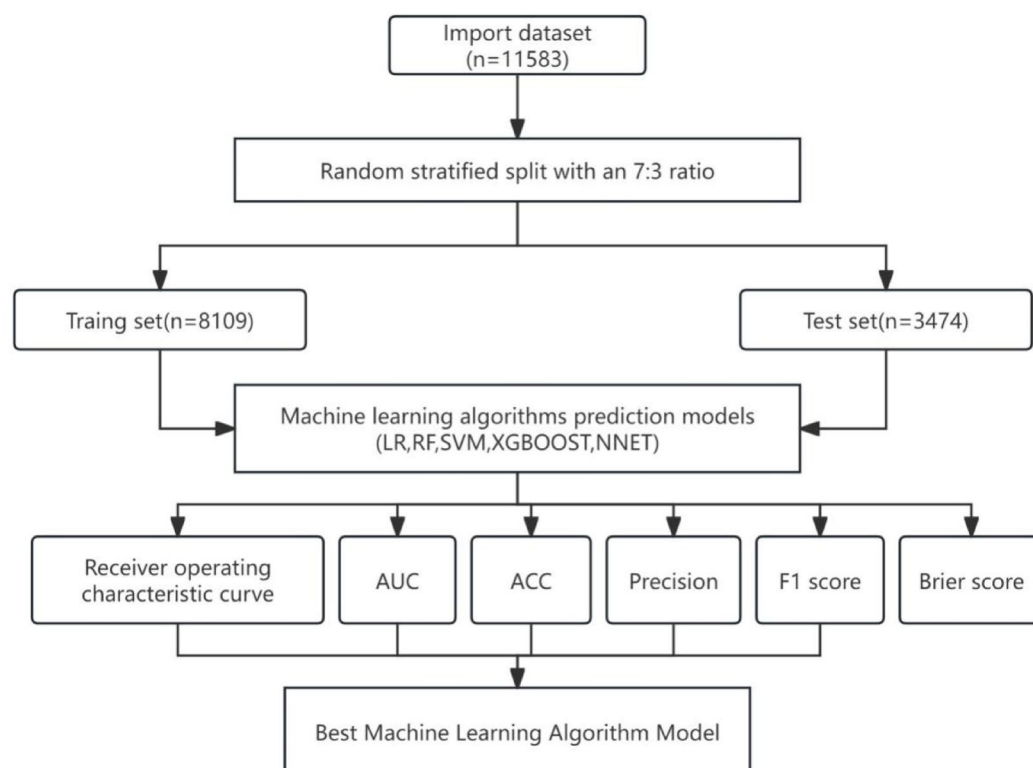


Fig. 1. Flowchart of patients through the study. LR, logistic regression; Nnet, neural network; RF, random forest; SVM, support vector machine; XGBoost, eXtreme Gradient Boosting; ACC, accuracy; AUC, area under the receiver operator characteristic curve.

	AUC	ACC	Precision	Recall	F1-score	Brier score
RF	0.7394	0.8938	0.9155	0.9727	0.9727	0.0866
XGBOOST	0.7460	0.8328	0.9327	0.8791	0.8791	0.6445
LR	0.7807	0.7104	0.9613	0.7094	0.7094	0.4441
SVM	0.7841	0.7009	0.9669	0.6942	0.6942	0.1847
NNET	0.6957	0.7110	0.9559	0.7145	0.7145	0.5135

Table 2. Evaluation of models. ACC, accuracy; AUC, area under the receiver operator characteristic curve.

(ROC) curves (AUC) values of RF, XGBoost, LR, SVM, and NNET in the test set were 0.7394, 0.7460, 0.7807, 0.7841, and 0.6957, respectively (Table 2). Table 2 shows that the SVM model outperformed the RF, XGBoost, and NNET models in terms of discrimination. The LR model had similarly good discrimination but was significantly less calibrated than the SVM model (Brier score: 0.4441 vs. 0.1847). The RF model outperformed the XGBoost and SVM models in terms of discrimination. Among the five ML models, the Brier scores of the two prediction sets (RF and SVM) were < 0.25. The SVM had a higher AUC (0.7841); better precision (0.9669); a Brier score of 0.1847; and lower accuracy (0.7009), recall (0.6942), and F1 (0.8091) scores than the RF model. Similarly, the Random Forest model also outperforms the LR. The ROC curves for the models in Fig. 2 show that the SVM had the largest AUC. We visualized the relationships between the important features (klotho) and CHD by fitting the restricted cubic splines (RCS) into a binary LR and then assessing the potential nonlinear relationships. Figure 3 shows that klotho did not have significant nonlinearity, thus suggesting a mainly negative linear relationship with the dependent variable. Figure 3 shows that the cutoff was 787.2 (odds ratio: 1), and the risk of CHD increased when the value was < 787.2. Figure 4a presents the restricted cubic spline (RCS) analysis of the relationship between age and Klotho levels in the coronary heart disease (CHD) group. The model demonstrated no significant nonlinear relationship ($p > 0.01$), with Klotho levels showing an overall declining trend as age increased. Figure 4b illustrates the RCS analysis for the non-CHD group, revealing a statistically significant nonlinear association ($p < 0.01$) between age and Klotho levels. A distinct inflection point was identified, suggesting a threshold effect. Notably, Klotho levels exhibited a marked acceleration in decline beyond the age of 65, highlighting a critical age-dependent reduction in this population. Figure 5 shows the importance of ranking the variables in the RF model. Additionally, we conducted a SHAP (SHapley Additive

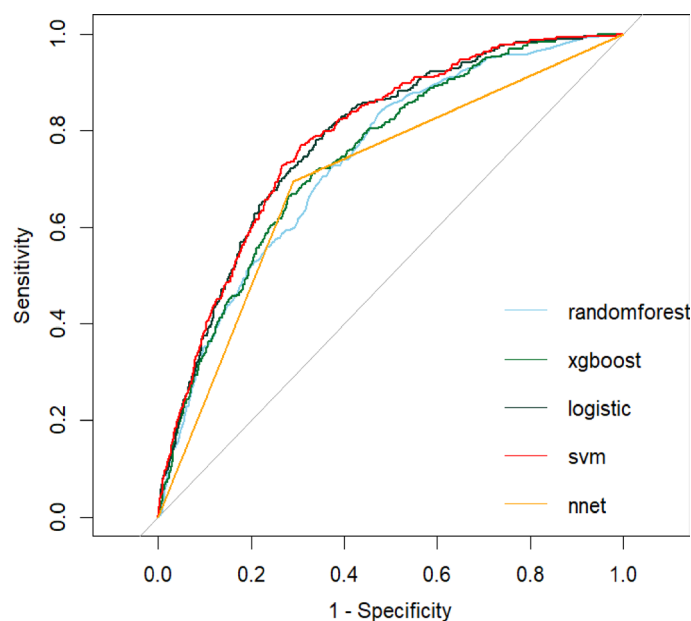


Fig. 2. Receiver operator characteristic curves of the five models.

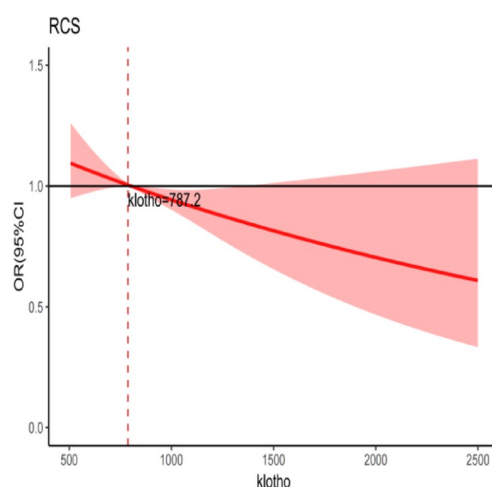


Fig. 3. Analysis of klotho and CHD using RCS. CHD, coronary heart disease; RCS, restricted cubic splines.

exPlanations) analysis to elucidate feature contributions in the RF model, revealing key drivers of its predictive performance and enhancing the interpretability of complex decision pathways (Figure S2).

Discussion

The number of deaths due to heart disease peaked during the mid-1960s, followed by a marked decline that continues today⁹. Nonetheless, CHD remains a major public health challenge that contributes to global disease burden. Despite improvements in preventive methods and treatments, CHD still imposes immense health and economic burdens in the United States and around the world¹⁰. Improvements at the diagnostic level can reduce this prevalence.

Klotho is an aging-suppressor gene¹¹, which has been found to predict the risk of developing cardiovascular disease^{12,13}. Atherosclerosis is the key pathological factor of coronary heart disease. According to Wang K et al. in their study^{14–17}, klotho is regarded as a cardiovascular protective factor because of its anti-atherosclerosis, anti-inflammatory and antioxidant effects, thus providing more theoretical basis for the prevention and treatment of coronary heart disease. Serum klotho concentrations display a consistent negative association with cardiovascular disease^{18,19}. In recent years, the method of combining nanomaterials to deliver klotho protein genes has been used as a treatment for other diseases associated with klotho. The study proves that upregulation of klotho gene may be a possible treatment and prevention method for aging-related complications^{20–22}, and it is likely to be used in the cardiovascular field in the future. Although supplementing Klotho protein holds potential

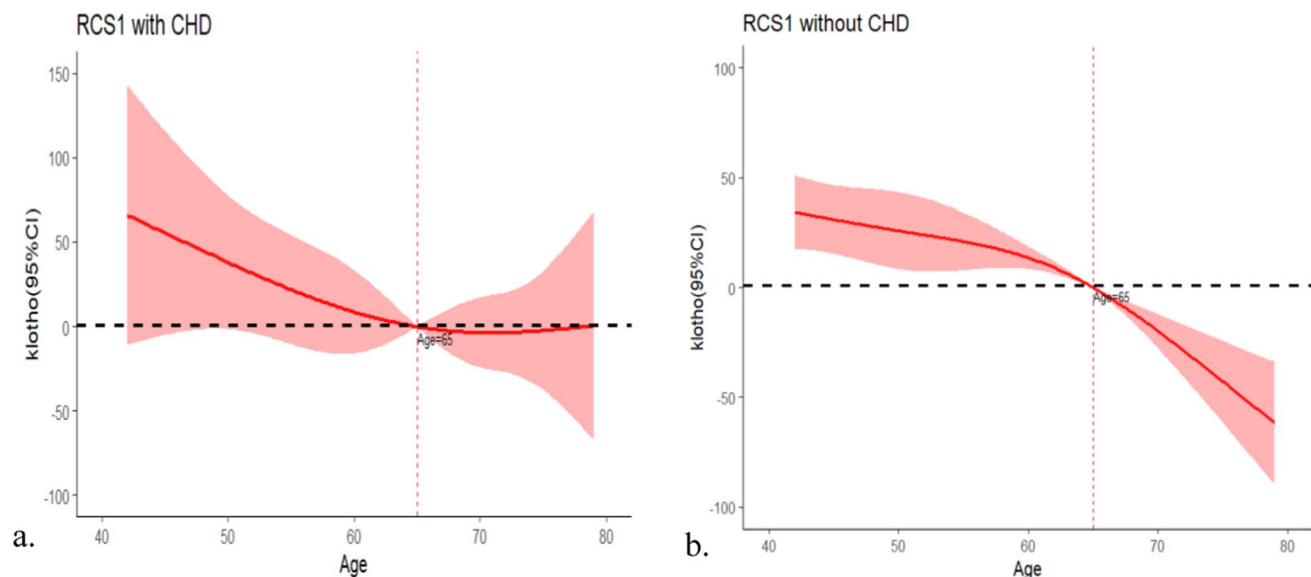


Fig. 4. Analysis of Age and klotho using RCS. CHD, coronary heart disease; RCS, restricted cubic splines.

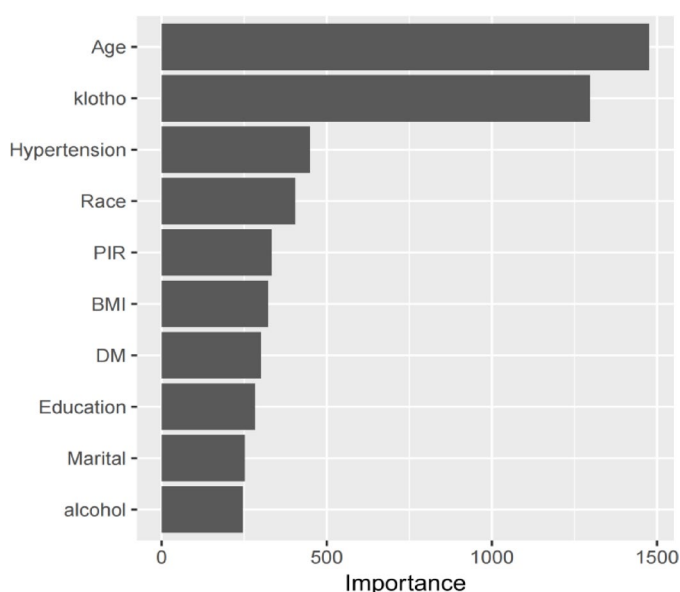


Fig. 5. Important Features for the Prediction of CHD. CHD, coronary heart disease; PIR, ratio of family income to poverty; BMI, Body Mass Index; DM, diabetes mellitus.

therapeutic value, its side effects and safety issues still require in-depth research. For example, exogenous Klotho protein might be recognized by the immune system as a foreign substance, triggering immune responses that could lead to allergies or inflammation. Additionally, Klotho protein interacts with FGF23²³, and excessive levels could disrupt the balance of phosphate and calcium metabolism. There are ongoing studies exploring the use of nanoparticle technology to deliver Klotho protein precisely to target tissues^{20–22}, but these remain at the animal research stage, and further studies are needed before clinical application can be realized. Future studies should focus on targeted delivery, dose optimization, and clinical trials to ensure its safety and efficacy.

ML is a branch of artificial intelligence (AI) that focuses on how computers learn patterns and rules from data to predict classify, decide, and other tasks. Among AI tools, ML is poised to potentially extend and augment the effectiveness of clinicians and revolutionize patient care²⁴. ML algorithms are extremely useful for predicting cardiovascular diseases²⁵ and have previously been combined with single or serial high-sensitivity cardiac troponin values to determine the probability of acute myocardial infarction²⁶.

To our knowledge, an ML algorithm to predict CHD based on klotho levels has not yet been created. We used NHANES dataset with large data volume and rich clinical information, and combined with the characteristics of machine learning to process big data to increase the reliability of the model. We investigated data about

indicators collected between 2007 and 2016 in National Health and Nutrition Examination Survey (NHANES) and used them as ML features to construct models.

When using klotho as a biomarker, machine learning (ML) models for CHD has higher accuracy in prediction compared to conventional statistical techniques. Random Forest is a powerful and versatile machine learning algorithm that leverages the strength of multiple decision trees to produce accurate and robust²⁷. The AUC showed that the model had good discrimination performance. In clinical settings, prioritize recall, F1 score, or Brier score over AUC alone. A low F1 score indicates that the model fails to balance precision and recall. A model with high AUC but low recall and F1 score may pose a risk of missed diagnoses in clinical applications, affecting its reliability and practicality. Further, the Brier score of the RF model was < 0.25 , thus indicating good predictive accuracy. The RF model performed better than the LR model, except for the AUC and precision. Therefore, considering the six evaluation indexes comprehensively, RF is the optimal model. Figure 5 shows the importance ranking of different features in building the RF model. Furthermore, the top three features in the RF tree model were age, klotho, and hypertension (Fig. 5). Klotho was second only to age. This indicated that klotho plays an important role in the construction of ML models and prediction of CHD onset. Our restricted cubic spline (RCS) analysis revealed that the age-dependent decline in klotho levels among coronary heart disease (CHD) patients aged over 65 was slower compared to non-CHD individuals (Fig. 4). This finding aligns with the conclusions of PAtlanta RS et al., whose study reported higher serum Klotho concentrations in older adults with a history of myocardial infarction²⁸, suggesting that disease status may decelerate age-related Klotho depletion. Whether this reflects disease-related compensatory mechanisms or is influenced by pharmacological interventions or comorbidities prevalent in older adults requires further investigation. Nevertheless, targeted monitoring of Klotho levels in high-risk age groups and exploring therapeutic time windows for Klotho replacement therapy hold clinical relevance. The discrepancy in feature importance rankings between Fig. 5 (based on Gini index) and Figure S2 (based on SHAP values) stems from fundamental differences in their evaluation perspectives and underlying logic. The Gini importance metric primarily reflects features' contributions to improving classification accuracy during node splitting in tree-based models, whereas SHAP values quantify the direct and localized impact of features on individual model predictions, accounting for both main effects and interactions. Regarding the discrepancy observed in the ROC curves of the NNET model compared to other models in Fig. 2, we attribute this to the neural network's heightened sensitivity to class imbalance. Despite implementing balancing strategies (e.g., SMOTE oversampling), its generalization capability may still slightly underperform ensemble methods such as random forests in handling minority-class predictions.

This study has some limitations. Most predictors included in this study were based on self-reports by participants, which could be subject to cognitive deficits and/or recall bias. We divided the entire dataset into training and validation sets for internal authentication, but we did not have external cohort studies to validate the performance of the prediction tool. The study results should be validated using external datasets, particularly those from other continents. The establishment of web is more convenient for clinical application and visualization of black-box machine learning models. We did not develop a web-based tool or mobile application on the basis of feature importance. We plan to address these issues in the future.

In conclusion, ML models based on serum klotho levels can predict CHD onset. Among the five ML models, RF showed the best predictive performance. We will continue to conduct external verification in subsequent studies.

Methods

Data sources and study population

This cross-sectional study analyzed nationally representative data from the US National Center for Health Statistics NHANES, which has continued in two-year cycles since 1999 to monitor the health and nutritional status of the US population. All NHANES protocols were approved by the National Center for Health Statistics Ethics Review Board, and written informed consent was obtained from participants in NHANES. Informed consent was not required for the present study because the information was downloaded from NHANES, which is an open secondary data source. As far as we are aware, no patients were involved in the planning, selection, or execution of NHANES.

Data from 50,588 persons who participated in five cycles of NHANES from 2007 to 2008 and 2015 to 2016 were selected. Among them, those aged < 40 years ($n = 31,244$) or had no serum klotho values ($n = 5580$) or clinically relevant indicators ($n = 2181$) were excluded from the analysis. Thus, we analyzed data from 11,583 persons.

Measurement of serum klotho

Samples on dry ice were inspected by personnel at the receiving area of the laboratory and then scanned. The data were compared with information on an accompanying electronic manifest and entered into the laboratory information system. The samples were stored at -80°C until defined batches were provided to technicians daily for analyses.

Outcome ascertainment

We identified CHD in questionnaire data reported by participants in NHANES. Experienced physicians subsequently diagnosed CHD using codes I00-09, I11, I13, and I20-51, which were determined according to the International Statistical Classification of Diseases and Related Health Problems (10th revision).

Feature selection and ML

Several features pooled in the model comprised clinically relevant indicators and serum klotho levels. These relevant features were driven to the model on the basis of domain knowledge and a literature search to obtain

an impression of predictors associated with major adverse cardiovascular events that were determined by traditional statistics. We used the ML algorithms LR, NNet, RF, SVM, and XGBoost to construct a classification model (see Supplemental Information for detailed procedure).

Statistical analysis

Non-normally distributed continuous and categorical variables are respectively presented as medians with interquartile ranges and as ratios (%). Continuous and categorical variables were analyzed using the Mann–Whitney U test and chi-square test, respectively.

We evaluated ML model discrimination by calculating the AUCs of sensitivity vs. Brier scores, specificity, accuracy, precision, recall, and F1. The AUC is a concordance index or C-statistic that reflects the probability that a risk score would be higher for a randomly selected person who experienced an event than one who did not. It is measured on a scale from 0.5 (a “coin flip”) to 1 (perfect discrimination)²⁹. The ROC curve was drawn from true and false positive rates on y and x axes, respectively. Brier scores combine the differentiation and calibration of the model and was used to evaluate the overall performance of the model; scores closer to zero indicated that the predicted values were closer to the actual values³⁰. The F1 score is an ML parameter that determines the performance of a model by combining its precision and recall scores. The following formula shows how the metrics were calculated:

$$\begin{aligned}\text{Accuracy} &= (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN}) \\ \text{Precision} &= \text{TP} / (\text{TP} + \text{FP}) \\ \text{Recall} &= \text{TP} / (\text{TP} + \text{FN}) \\ \text{F1 score} &= (2 * \text{precision} * \text{recall}) / (\text{precision} + \text{recall})\end{aligned}$$

RCSs effectively capture nonlinear relationships in data through piecewise fitting and smooth transitions, while also reducing the risk of overfitting. Their flexibility and broad applicability make them a powerful tool for handling nonlinear data. RCSs are used to model the relationship between important continuous variable and CHD, particularly in regression analysis, because it allows the capture of nonlinear relationships and avoidance of overfitting while maintaining smoothness.

All data were statistically analyzed using SPSSAU software (SPSSAU, Haidian, China) or R 4.4.0 (R Foundation for Statistical Computing, Vienna, Austria).

Data availability

Data supporting the study findings are publicly available. The full lists of the records identified through the database search are available upon reasonable request from the corresponding author (<http://www.cdc.gov/nchs/nhanes.htm>).

Received: 19 July 2024; Accepted: 19 May 2025

Published online: 27 May 2025

References

- Hasbani, N. R. et al. American Heart Association's life's simple 7: Lifestyle recommendations, polygenic risk, and lifetime risk of coronary heart disease. *Circulation* **145**, 808–818 (2022).
- Khoja, A. et al. Modifiable and non-modifiable risk factors for premature coronary heart disease (PCHD): Systematic review and meta-analysis. *Heart Lung Circ.* **33**, 265–280 (2024).
- Xu, J. P. et al. Associations between serum soluble α-klotho and the prevalence of specific cardiovascular disease. *Front. Cardiovasc. Med.* **9**, 899307. <https://doi.org/10.3389/fcvm.2022.899307> (2022).
- Yang, Z. et al. The prognostic value of serum α-klotho in age-related diseases among the US population: A prospective population-based cohort study. *Prev. Med. Rep.* **42**, 102730. <https://doi.org/10.1016/j.pmedr.2024.102730> (2024).
- Tyurenkov, I. N., Perfilova, V. N., Nesterova, A. A. & Glinka, Y. Klotho protein and cardio-vascular system. *Biochemistry (Mosc)* **86**, 132–145 (2021).
- Bi, X., Yang, K., Zhang, B. & Zhao, J. The protective role of klotho in CKD-associated cardiovascular disease. *Kidney Dis. (Basel)* **6**, 395–406 (2020).
- Greener, J. G., Kandathil, S. M., Moffat, L. & Jones, D. T. A guide to machine learning for biologists. *Nat. Rev. Mol. Cell Biol.* **23**, 40–55 (2022).
- Dritsas, E. & Trigka, M. Data-driven machine-learning methods for diabetes risk prediction. *Sensors (Basel)* **22**, 5304. <https://doi.org/10.3390/s22145304> (2022).
- Dalen, J. E., Alpert, J. S., Goldberg, R. J. & Weinstein, R. S. The epidemic of the 20(th) century: Coronary heart disease. *Am. J. Med.* **127**, 807–812 (2014).
- Virani, S. S. et al. Heart disease and stroke statistics-2021 update: A report from the American Heart Association. *Circulation* **143**, e254–e743 (2021).
- Xu, Y. & Sun, Z. Molecular basis of Klotho: From gene to function in aging. *Endocr. Rev.* **36**, 174–193 (2015).
- Semba, R. D. et al. Plasma klotho and cardiovascular disease in adults. *J. Am. Geriatr. Soc.* **59**, 1596–1601 (2011).
- Chen, Z. et al. Serum klotho modifies the associations of 25-hydroxy vitamin d with all-cause and cardiovascular mortality. *J. Clin. Endocrinol. Metab.* **109**, 581–591 (2024).
- Bi, J. et al. Relationships of serum FGF23 and α-klotho with atherosclerosis in patients with type 2 diabetes mellitus. *Cardiovasc. Diabetol.* **23**(1), 128. <https://doi.org/10.1186/s12933-024-02205-2> (2024).
- Ajay, A. K. et al. Local vascular Klotho mediates diabetes-induced atherosclerosis via ERK1/2 and PI3-kinase-dependent signaling pathways. *Atherosclerosis* **396**, 118531. <https://doi.org/10.1016/j.atherosclerosis.2024.118531> (2024).
- Yao, Y., Wang, Y., Zhang, Y. & Liu, C. Klotho ameliorates oxidized low density lipoprotein (ox-LDL)-induced oxidative stress via regulating LOX-1 and PI3K/Akt/eNOS pathways. *Lipids Health Dis.* **16**(1), 77. <https://doi.org/10.1186/s12944-017-0447-0> (2017).
- Nagai, R. et al. Endothelial dysfunction in the klotho mouse and downregulation of klotho gene expression in various animal models of vascular and metabolic diseases. *Cell Mol. Life Sci.* **57**, 738–746. <https://doi.org/10.1007/s000180050038> (2000).
- Lee, J. et al. Association between serum klotho levels and cardiovascular disease risk factors in older adults. *BMC Cardiovasc. Disord.* **22**, 442. <https://doi.org/10.1186/s12872-022-02885-2> (2022).

19. Cai, J. et al. Association between serum Klotho concentration and heart failure in adults, a cross-sectional study from NHANES 2007–2016. *Int. J. Cardiol.* **370**, 236–243 (2023).
20. Li, H. et al. Nanoparticle-mediated Klotho gene therapy prevents acute kidney injury to chronic kidney disease transition through regulating PPAR α signaling in renal tubular epithelial cells. *Biomaterials* **315**, 122926. <https://doi.org/10.1016/j.biomaterials.2024.122926> (2025).
21. Wang, P. et al. Attenuation of osteoarthritis progression via locoregional delivery of Klotho-expressing plasmid DNA and Tanshinon IIA through a stem cell-homing hydrogel. *J. Nanobiotechnol.* **22**(1), 325. <https://doi.org/10.1186/s12951-024-02608-z> (2024).
22. Li, C. et al. An in situ sustained-release chitosan hydrogel to attenuate renal fibrosis by retaining Klotho expression. *Biomater. Res.* **24**(28), 0099. <https://doi.org/10.34133/bmr.0099> (2024).
23. Belfiori, M., Lazzari, L., Hezzell, M., Angelini, G. D. & Dong, T. Transcriptomics, proteomics and bioinformatics in atrial fibrillation: A descriptive review. *Bioengineering (Basel)* **12**(2), 149. <https://doi.org/10.3390/bioengineering12020149> (2025).
24. Itchhaporia, D. Artificial intelligence in cardiology. *Trends Cardiovasc. Med.* **32**, 34–41 (2022).
25. Weng, S. F., Reps, J., Kai, J., Garibaldi, J. M. & Qureshi, N. Can machine-learning improve cardiovascular risk prediction using routine clinical data? *PLoS ONE* **12**, e0174944. <https://doi.org/10.1371/journal.pone.0174944> (2017).
26. Doudesis, D. et al. Machine learning for diagnosis of myocardial infarction using cardiac troponin concentrations. *Nat. Med.* **29**, 1201–1210 (2023).
27. Hu, J. & Szymczak, S. A review on longitudinal data analysis with random forest. *Brief Bioinform.* **24**(2), bbad002. <https://doi.org/10.1093/bib/bbad002> (2023).
28. Paula, R. S. et al. Serum Klotho (but not haplotypes) associate with the post-myocardial infarction status of older adults. *Clinics (Sao Paulo)* **71**(12), 725–732. [https://doi.org/10.6061/clinics/2016\(12\)09](https://doi.org/10.6061/clinics/2016(12)09) (2016).
29. Vickers, A. J. & Holland, F. Decision curve analysis to evaluate the clinical benefit of prediction models. *Spine J.* **21**, 1643–1648 (2021).
30. Redelmeier, D. A., Bloch, D. A. & Hickam, D. H. Assessing predictive accuracy: How to compare Brier scores. *J. Clin. Epidemiol.* **44**, 1141–1146 (1991).

Acknowledgements

We are grateful to all the investigators and staff members who participated in this study. We would also like to thank Editage for the English language editing. This study was supported by grants from the Lvliang Key Laboratory of Cardiovascular Disease (2020ZDSYS16) and the Cultivation Key Laboratory of Metabolic Cardiovascular Diseases Research (No. TSGJ001) and by the Key R&D Projects of Introducing High-Level Scientific and Technological Talents in Lvliang City (2019115).

Author contributions

YY and YZ: designed and drafted the study and interpreted the data. MM and LB: critically revised the manuscript for intellectual content. All authors were involved in the data collection and approved the final version of the manuscript for publication.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-03234-x>.

Correspondence and requests for materials should be addressed to M.M. or L.B.

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.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025