scientific reports

Check for updates

OPEN

Development and validation of a prediction model for ED using machine learning: according to NHANES 2001–2004

Xing-Yu Chen[1], Wen-Ting Lu[2], Di Zhang[3], Mo-Yao Tan[4] & Xin Qin[1]✉

Erectile Dysfunction (ED) is a form of sexual dysfunction in males that imposes significant health and financial burdens globally. Despite its high prevalence, diagnosing ED remains challenging due to the limitations of current diagnostic methods and patients' reluctance to seek medical help. Currently, some studies have used machine learning techniques for developing ED prediction models, but the performance and interpretability of existing models need to be further improved. This study utilized data from the National Health and Nutrition Examination Survey (NHANES) for the years 2001 to 2004, adhering to the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) statement. After excluding male respondents who did not meet the study criteria, a total of 3,869 participants were included. Gradient boosting decision tree (GBDT) algorithms (XGBoost, CatBoost, LightGBM) were used to develop the ED prediction model. Data preprocessing, feature selection, model evaluation, and interpretability analysis were performed to ensure the reliability and effectiveness of the model. The model evaluation results revealed that the AUC values are XGBoost: 0.887 ± 0.016; LightGBM: 0.879 ± 0.016; CatBoost: 0.871 ± 0.019. The F1-Scores are XGBoost: 0.695 ± 0.023; LightGBM: 0.681 ± 0.025; CatBoost: 0.681 ± 0.025. The Recall values are XGBoost: 0.789 ± 0.026; LightGBM: 0.739 ± 0.030; CatBoost: 0.711 ± 0.030. These results confirmed that the XGBoost model is the best-performing ED prediction model in this study. Interpretability analysis results of the XGBoost model showed that age, obesity, cardiovascular risk factors, prostate-related diseases, and socioeconomic status are key features for predicting ED, playing a significant role in the ED mechanism. Therefore, we believe the ED prediction model trained in this study has strong predictive performance and high interpretability. This model can help to expand the diagnostic options for ED, improve the diagnosis rate of ED, and assist doctors in early intervention for patients with ED, ultimately improving patient prognosis.

Keywords Erectile Dysfunction, Machine learning, XGBoost, National Health and Nutrition Examination Survey, Prediction model

Erectile dysfunction (ED) is defined as the persistent or repeated inability to attain and/or maintain penile erection sufficient for sexual satisfaction. It is estimated that around 150 million men worldwide suffer from ED, mainly affecting those over the aged 40 and above[1–3]. ED significantly reduces the quality of sexual life, adversely affecting mental health and family relationships[4,5]. Moreover, ED is closely associated with cardiovascular diseases (CVD) risk factors, including hypertension, high cholesterol, diabetes, and smoking habits[6]. The global prevalence of ED is gradually increasing as the number of men with cardiovascular risk factors increases and the population ages[7]. Authoritative studies have indicated that the pathogenesis of ED is congruent with CVD, early detection and treatment of ED can aid in effectively reducing future CVD risks[8,9], and societal healthcare costs[10].

However, both patients and doctors do not have sufficient understanding of ED. On the one hand, due to the sensitive nature of ED, patients often exhibit reluctance to seek medical advice, leading to suboptimal diagnosis rates[11,12]. On the other hand, medical professionals have not fully recognized the importance of ED, and ED's assessment is often overlooked, hindering the diagnosis and treatment of ED and related diseases[13–15].

[1]Chengdu Integrated TCM and Western Medicine Hospital, Chengdu, Sichuan, China. [2]XinDu Hospital of Traditional Chinese Medicine, Chengdu, Sichuan, China. [3]West China School of Pharmacy, Sichuan University, Chengdu, Sichuan, China. [4]Chengdu University of Traditional Chinese Medicine, Chengdu, Sichuan, China. ✉email: qinxin72157616@126.com

In addition, the diagnostic techniques commonly employed in clinical practice have their limitations. For instance, the RigiScan, which provides objective data on erection firmness, is expensive and time-consuming; while the penile color duplex Doppler ultrasound (CDDU) can assess penile vascular performance, it has a high false-positive rate[16,17]. These circumstances underscore the need for further improvement and development of diagnostic methods for ED. Currently, the swift progression of Artificial Intelligence (AI) technology has been instrumental in the medical field, where numerous studies have effectively harnessed machine learning, neural networks, and other methodologies for analyzing high-dimensional datasets, and developed variety of disease prediction models[18–20]. These studies demonstrate that prediction models constructed using machine learning techniques can enhance the accuracy of diagnosis and medical decisions[21]. In this study, we aimed to utilize data from the National Health and Nutrition Examination Survey (NHANES) for the years 2001 to 2004, using machine learning techniques to develop and validate a prediction model for ED, which seeks to enhance the diagnostic capabilities for ED.

Methods

Respondents and study design

This study adheres to the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) statement[22]. The data utilized in this research were derived from the National Health and Nutrition Examination Survey (NHANES), a project centered on public health. This health survey was initiated by the Centers for Disease Control and Prevention (CDC) and the National Center for Health Statistics (NCHS), conducted to gather and periodically release research data to the public at no cost. Written informed consent was obtained from all NHANES respondents or their proxies, and the entire data collection procedure for respondents, which includes anthropometric measurements, health and nutrition questionnaires, and laboratory tests, has been approved by the NCHS Research Ethics Review Board. NHANES conducted surveys on the condition of ED among respondents only between specific time periods (January 1, 2001, to December 31, 2004). Consequently, we extracted public data files from NHANES corresponding to the period to create our dataset. Within the time periods, a total of 21,161 respondents participated. Respondents were excluded from the analysis if they were female, under the age of 20, had missing ED questionnaire data, or responded with "refused" or "Don't know" to questions related to the included features and outcome variables. The final study population utilized for model training comprised 3,869 respondents.

ED Assessment

The outcome variable in this study was set as the incidence of erectile dysfunction (ED). Given that the condition of ED pertains to a sensitive health risk behavior topic, NHANES employed the Mobile Examination Center Audio-Computer-Assisted Self-Interview (MEC ACASI) system to assess the erectile dysfunction status among respondents. The specific approach was as follows: respondents were asked to answer questions in a private room in the MEC without the interviewer, respondents were allowed to hear questions through earphones or to read questions on a computer screen, then they can touch the computer screen to indicate their response at their own speed. The question is: many men experience problems with sexual intercourse. How would you describe your ability to get and keep an erection adequate for satisfactory intercourse? The answer options are: 1)always able or almost always able, 2) usually able, 3) sometimes able, 4) never able. This self-reported assessment method has been considered in previous studies to accurately predict clinically diagnosed ED[23]. Within the NHANES dataset, the distribution of respondents across different ED classification levels is uneven. To address the issue of data imbalance and enhance the predictive performance and practical utility of the model, this study opted to construct a binary classification model. Respondents who answered "sometimes able" or "never able" were classified as having ED, whereas those who responded "always able or almost always able" or "usually able" were classified as not having ED.

Assessment of covariates

The demographic, examination, laboratory and questionnaire data extracted from the NHANES datasets were identified as covariates, according to previous clinical practices and methodologies described in the literature. The covariates include: Demographic Data: Age in Months, Education Level, Ratio of family income to poverty threshold (PIR); Examination Data: Body Measurement (Body Mass Index (BMI), Arm Circumference, Waist Circumference, Standing Height), Cardiovascular fitness level; Laboratory Data: Cholesterol (HDL-Cholesterol), Complete Blood Count (Hemoglobin, Lymphocyte number, Segmented neutrophils number, Platelet count), Prostate specific antigen (PSA), Testosterone, Vitamins (b-cryptoxanthin, Combined Lutein/zeaxanthin, Vitamin B12, Vitamin D, Folate); Questionnaire Data: Alcohol Drinking, Smoking Status, Hypertension, High cholesterol, Cardiovascular Health (ever had pain or discomfort in the chest, shortness of breath on stairs/inclines), Diabetes, Cancer/Malignancy, Physical Activity (Hours watched TV past 30 days; Moderate activity over the past 30 days), Mental Health (Depression/Generalized Anxiety Disorder/Panic Disorder), Prostate Conditions (Usually have trouble trying to urinate, Infection of prostate, Prostate disease, Enlarged prostate, Ever had a PSA test), Sexual Behavior (Circumcised or not, Sexually transmitted diseases or not), Social Support (Spouse emotional support). Respondents who had smoked at least 100 cigarettes in life were assigned to "Yes" group in "Smoking"; who had at least 12 drinks of any type of alcoholic beverage in life were assigned to "Yes" group in "Alcohol Drinking"; who had taken prescription for hypertension were considered to have hypertension; who had been told to take prescription for cholesterol were considered to have high cholesterol; who had been told that they had diabetes by doctors were considered to have diabetes.

Additionally, this study incorporated the Systemic immune-inflammation index (SII) to assess the local immune response and systemic inflammation of the respondents[24]. The calculation formulas for these variables are: SII: (Platelet count × Neutrophils count)/Lymphocytes count.

A detailed description of how various covariates were measured and collected can be found on the NHANES website (https://wwwn.cdc.gov/nchs/nhanes).

Data preprocessing

Data preprocessing in this study was based on the NHANES analytic guidelines[25]. We preprocessed The NHANES data that met our screening criteria (3,869 respondents). First, the Interquartile Range (IQR) method was applied to detect and handle the continuous data's outliers. The values below the first quartile (Q1) − 1.5*IQR and above the third quartile (Q3) + 1.5*IQR were considered outliers and replaced with the nearest boundary values. The weighted comparisons between the ED group and the non-ED group were performed after handling outliers, the features showing no significant difference between the two groups were excluded. Logarithmic transformations were then applied to enhance the model's stability and adaptability[26]. The preprocessed dataset was split into train and test sets using stratified sampling with an 8:2 ratio.

The NHANES data utilized in this study contained missing values. The NHANES datasets were employed a complex multi-stage random sampling method, resulting in relatively small sample sizes for certain populations, such as certain age groups or people with specific health conditions. This leads to a higher incidence of missing values, especially in categorical variables. To enhance the richness and quality of the datasets and improve the accuracy of predictions, we employed the random forest method for imputing missing values. This method can handle missing data for both numerical and categorical variables and has been previously applied in research on prediction model[21]. To avoid data leakage, we employed the random forest imputation method separately on the train and test sets, and ensured that the target variable (ED) was excluded during the imputation.

Feature engineering plays an important role in the data preprocessing phase of machine learning. Effective feature engineering can reduce data dimensionality, improve model performance, and increase computational efficiency and accuracy[27]. The GBDT algorithms have advantages in handle missing values and processing categorical variables, making them suitable for feature selection. In this study, we fitted the XGBoost, LightGBM, and CatBoost models on the train set and selected the top 25 most important features from each model. The intersection of these features was taken as the selected features.

Development of machine learning models

Three types of GBDT algorithms were selected to construct the ED prediction model: XGBoost, CatBoost, and LightGBM. These algorithms can effectively manage the non-linear relationships between features, boasting considerable computational power and generalization ability. We trained the prediction models using the features data that had been filtered through the preprocessing steps. This study used the foreach and parallel package in R for hyperparameter tuning of the models. Initially, we created a parallel computing cluster and defined hyperparameter grids for the models. Then, we iterated over the hyperparameter grids, and performed 5-fold cross-validation to evaluate the AUC values of the models under different combinations of parameters within the grid, determining the optimal parameters for the training models. The optimized parameters are displayed in the supplementary (Supplementary Table S2).

A model prediction probability greater than 0.5 was designated as the positive group (ED group). Subsequently, stratified 5-fold cross-validation was applied as the model training method. This technique involves dividing the selected train dataset (80%) into 5 equal-sized folds while maintaining the proportion of each category of samples in each fold consistent with that in the original dataset. One of the folds was randomly selected as the validation set (16%) to verify the model training outcome, and the remaining 4 folds were merged to form the model training set (64%). This training and validation process was repeated 5 times, with the validation and training sets being reselected before each iteration. Based on the validation set AUC value, we selected the best-performing fold from each algorithms as the final models, which were then evaluated and compared on the test set (The test set (20%) had been adjusted based on the feature selection results of train set (80%)). In the test set, AUC value, confusion matrix, calibration curves, and decision curve analysis (DCA) were used to evaluate the model's discriminative ability, accuracy, and clinical utility. To enhance the reliability of the model performance comparison results, a 1000-time bootstrap was applied to compute the means and standard deviations (Mean ± SD) of each metric for model comparison. Upon identifying the best-performing model through these metrics, the interpretability analysis was conducted using Feature Importance, SHAP (Shapley additive explanations) summary plots, and SHAP dependence plots.

To strengthen the robustness of the results, we conducted multiple sensitivity analyses on the best-performing model to clarity the effects of various data preprocessing methods on model performance. We analyzed the effects of logarithmic transformations, imputation techniques and feature selection methods on model performance. Due to the imbalance between the number of ED and non-ED respondents in dataset, we also analyzed the effect of employing the Borderline-Synthetic Minority Over-sampling Technique (Borderline-SMOTE) and Edited Nearest Neighbors (ENN) to oversample the model training dataset (64%) in each iteration before training commenced on model performance.

Statistical analysis

Following the NHANES analytic guidelines, we employed sampling weights, Primary Sampling Unit (PSU), and stratum data to conduct weighted comparisons between the ED group and the non-ED group within the NHANES dataset by using the survey package in R. Continuous variables were represented by means and standard deviations (Mean ± SD), and comparisons were made using a Design-based T-test. Categorical

variables were presented as counts (percentages), and comparisons were conducted using Pearson's chi-squared test with Rao & Scott adjustment. *P*-values were two-sided and $P < 0.05$ was considered significant.

In order to compare the performance of three models, the first step is to select the metrics for analysis. The Shapiro-Wilk test was then performed on the data corresponding to these metrics for each model to determine whether they followed a normal distribution. Levene's test was then applied to determine whether there were significant differences in data variances for the selected metrics between models. If the data are normally distributed and variances are equal, Fisher's ANOVA test was performed for inter-group comparisons, followed by Tukey's Honestly Significant Difference (Tukey HSD) post-hoc analysis. If the data are normally distributed but variances are not equal, Welch's ANOVA test was performed, followed by the Games-Howell post-hoc analysis. If neither assumption is met, the Kruskal-Wallis test and Dunn's post-hoc analysis were proceeded.

For the pairwise comparisons in the sensitivity analysis, Welch's t-test was used if the data are normally distributed and variances are equal. If the data do not follow a normal distribution, the Wilcoxon rank-sum test was conducted. For the comparisons among three groups in the sensitivity analysis, the statistical methods followed the same method of model performance comparison. The performance metrics of the models were presented as Mean ± SD.

Data processes, model construction, and statistical analyses were performed using R software, version 4.3.2. Additionally, we also created a flowchart to make the research process clearer (Fig. 1).

Results

Baseline characteristics of respondents

A total of 3869 respondents were included in this study, and 49 variables were selected as preliminary features. Group analysis was conducted based on whether ED was present or not (Table 1).

There were significant differences between the ED group and the NON-ED group in terms of age, PIR, education level, diabetes, hypertension, high cholesterol, and PSA levels ($P < 0.01$). Compared to respondents without ED, those with ED tended to be older, have a higher probability of hypertension and high cholesterol, have higher levels of BMI and SII, and more likely to have prostate-related diseases. Moreover, they generally had lower education levels and PIR. No significant differences were found between the two groups of respondents regarding vitamin B12, HDL, alcohol drinking, cardiovascular fitness level, sexually transmitted diseases (STD) and other features. The 12 features showing no significant difference between the two groups were excluded.

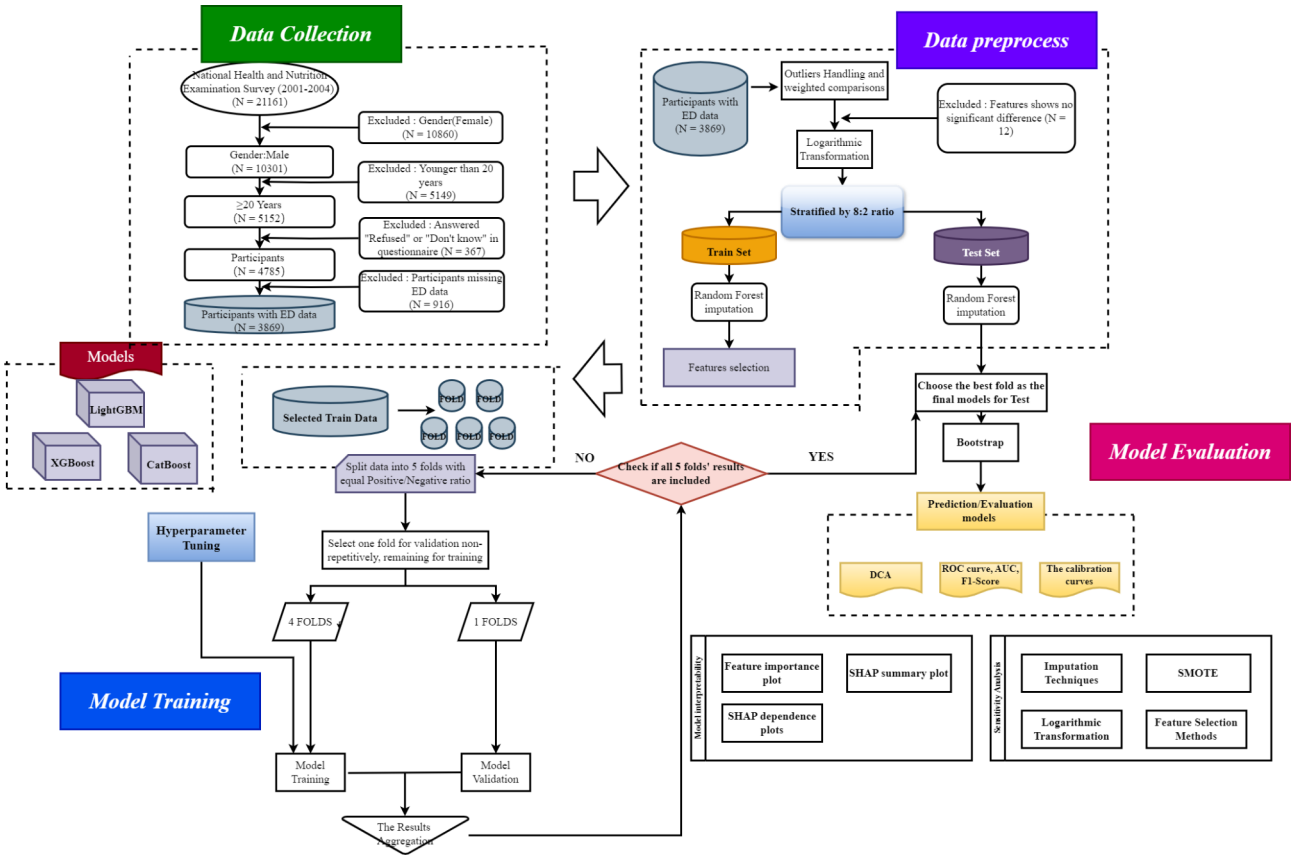

Fig. 1. Flow chart of data collection, data preprocessing, and model development.

Characteristics	Value	ED	NON-ED	*P*-value
N		1105(28.6%)	2764(71.4%)	
Age (Months)	240–1016	775.99 ± 178.46	520.05 ± 183.7	< 0.01
Family PIR	0–5	2.56 ± 1.56	2.94 ± 1.62	< 0.01
BMI (kg/m2)	15.41–39.82	28.24 ± 5.06	27.69 ± 4.88	< 0.01
Waist Circumference (cm)	63.7-135.7	103.54 ± 13.81	98.31 ± 13.76	< 0.01
Arm Circumference (cm)	22.85–44.05	32.82 ± 4.31	33.94 ± 4.02	< 0.01
Height (cm)	154.8-195.6	173.09 ± 7.38	175.94 ± 7.61	< 0.01
Testostosterone (ng/mL)	0.06–10.3	4.12 ± 2.03	5.31 ± 2.02	< 0.01
Beta-cryptoxanthin (ug/dL)	0.14-22.256	8.41 ± 5.5	9.06 ± 5.61	< 0.01
Folate, serum (nmol/L)	2.5–61.3	32.92 ± 15.28	26.8 ± 11.84	< 0.01
Vitamin D (nmol/L)	9.1-115.65	58.03 ± 19.95	59.53 ± 20.56	< 0.01
Age at first BPH diagnosis	27.5–87.5	60.67 ± 12.38	54.44 ± 12.67	< 0.01
PSA. total (ng/mL)	0.1–3.85	1.68 ± 1.25	1.23 ± 0.98	< 0.01
Lymphocyte number	0.5–3.75	1.88 ± 0.66	2.07 ± 0.65	< 0.01
Hemoglobin (g/dL)	12.25–18.25	14.8 ± 1.31	15.4 ± 1.11	< 0.01
Segmented neutrophils number	0.6–7.85	4.28 ± 1.44	4.1 ± 1.48	< 0.05
Platelet count (%)	91–403	238.46 ± 62.26	254.42 ± 58.47	< 0.01
Systemic immune-inflammation index	12.61-1174.65	578.58 ± 279.61	531,79 ± 249.52	< 0.01
Age when first had sexual intercourse	10.5–22.5	16.6 ± 3.01	16.6 ± 2.93	0.5697
Women sex intercourse partners/life	0-45.5	12.53 ± 13.34	12.84 ± 13.25	0.9326
Direct HDL-Cholesterol (mg/dL)	19-76.5	47.04 ± 11.88	47.33 ± 12.16	0.3049
Combined Lutein/zeaxanthin (ug/dL)	1.7-33.955	15.75 ± 7.32	15.9 ± 7.05	0.668
Vitamin B12, serum (pmol/L)	81.18-698.978	365.67 ± 150.43	360.8 ± 130.69	0.2427
Education				< 0.01
Less Than High School	1	424(38.4%)	614(22.2%)	
High School Diploma	2	235(21.3%)	742(26.8%)	
More Than High School	3	446(40.4%)	1408(50.9%)	
Smoking				< 0.01
Yes	1	762(69%)	1531(55.4%)	
No	2	343(31%)	1233(44.6%)	
Hours watch TV or videos past 30 days				< 0.01
Less than 1 h	0	51(8.8%)	173(13.3%)	
1 h	1	60(10.3%)	214(16.4%)	
2 h	2	143(24.7%)	405(31%)	
3 h	3	124(21.4%)	198(15.2%)	
4 h	4	67(11.6%)	130(10%)	
5 h or more	5	128(22.1%)	167(12.8%)	
none	6	7(1.2%)	18(1.4%)	
Moderate activity over past 30 days				< 0.01
Yes	1	470(42.5%)	1440(52.1%)	
No	2	580(52.5%)	1297(46.9%)	
Unable to do activity	3	55(5%)	27(1%)	
Hypertention				< 0.01
Yes	1	522(90.2%)	427(69%)	
No	2	57(9.8%)	192(31%)	
High Cholesterol				< 0.01
Yes	1	337(73.6%)	303(47.4%)	
No	2	121(26.4%)	336(52.6%)	
Diabetes				< 0.01
Yes	1	256(23.2%)	140(5.1%)	
No	2	849(76.8%)	2624(94.9%)	
Cancer/Malignancy				< 0.01
Yes	1	239(21.6%)	128(4.6%)	
No	2	866(78.4%)	2636(95.4%)	
Trouble trying to urinate				< 0.01
Continued				
Yes	1	174(17.1%)	73(4.8%)	
No	2	846(82.9%)	1449(95.2%)	
Empty bladder after urinating				< 0.01
Yes	1	805(78.9%)	1332(87.5%)	

Characteristics	Value	ED	NON-ED	*P*-value
No	2	215(21.1%)	190(12.5%)	
Prostate Disease				< 0.01
Yes	1	353(31.9%)	202(7.3%)	
No	2	752(68.1%)	2562(92.7%)	
Enlarged prostate				< 0.01
Yes	1	321(31.5%)	172(11.3%)	
No	2	699(68.5%)	1350(88.7%)	
Enlargement was BPH				< 0.01
Yes	1	194(60.4%)	131(76.2%)	
No	2	127(39.6%)	41(23.8%)	
Enlargement due to cancer				< 0.01
Yes	1	66(52%)	6(14.6%)	
No	2	61(48%)	35(85.4%)	
Ever had a PSA test				< 0.01
Yes	1	595(58.3%)	553(36.3%)	
No	2	425(41.7%)	969(63.7%)	
Ever had a rectal exam				< 0.01
Yes	1	877(86%)	1209(79.4%)	
No	2	143(14%)	313(20.6%)	
Circumcised or not				< 0.01
Yes	1	170(59.4%)	1698(74.4%)	
No	2	116(40.6%)	584(25.6%)	
Pain or discomfort in chest				< 0.01
Yes	1	375(36.8%)	423(27.8%)	
No	2	645(63.2%)	1099(72.2%)	
Shortness of breath on stairs/inclines				< 0.01
Yes	**1**	457(44.8%)	72(24.4%)	
No	**2**	563(55.2%)	1150(75.6%)	
Prostate Exam				< 0.01
Yes	**1**	899(88.1%)	1226(80.6%)	
No	**2**	121(11.9%)	296(19.4%)	
Alcohol Drinking				
Yes	1	137(63.1%)	256(57.1%)	0.06179
No	2	80(36.9%)	192(42.9%)	
Prostate Inflammation				
Yes	1	16(1.7%)	13(0.9%)	0.5783
No	2	939(98.3%)	1455(99.1%)	
STD				0.7219
Yes	1	15(5.6%)	131(5.9%)	
No	2	254(94.4%)	2080(94.1%)	
Cardiovascular fitness level				0.108
Low	1	10(18.5%)	151(14.3%)	
Moderate	2	21(38.9%)	369(34.9%)	
High	3	23(42.6%)	536(50.8%)	
Genital Herpes				0.4633
Yes	1	8(3%)	52(2.4%)	
No	2	261(97%)	2160(97.6%)	
Genital Warts				0.8761
Yes	1	7(2.6%)	70(3.2%)	
No	2	262(97.4%)	2142(96.8%)	
Gonorrhea				0.3328
Yes	1	NA(NA%)	10(0.5%)	
No	2	269(100%)	2202(99.5%)	

Table 1. Basic characteristics of respondents ($n = 3869$) in the NHANES 2001–2004. Mean ± SD was for continuous variables, N(%) was for categorical variables; $P < 0.05$ indicating a significant difference between the two groups. BMI, Body mass index; BPH, Benign prostatic hyperplasia; PIR, Ratio of family income to poverty threshold; PSA, Prostate-Specific Antigen; SII, Systemic immune-inflammation index; STD, Sexually transmitted diseases.

Features selection

Using the GBDT's feature importance rank for further selection among the 37 features. 23 features were selected for model training. These include: Demographic Characteristics (Age, Education Level, PIR), Body Measurements (Standing Height, Waist Circumference, Arm Circumference, Body Mass Index), Disease Status (Diabetes, Cardiovascular Condition, Cancer, Prostate Disease), Physical Activity (Time of watching videos, Moderate activity condition)Laboratory Indicators (PSA, Folate, Vitamin D, Testosterone, Hemoglobin, Lymphocytes, Platelet, neutrophils, SII).

Model performance comparison

ED prediction models were trained using selected features on the train set (80%), and model performance was evaluated on the test set (20%) (Supplementary Table S1, Supplementary Figure S1). The AUC values on the test set of XGBoost was 0.887 ± 0.016; LightGBM was 0.879 ± 0.016; CatBoost was 0.871 ± 0.019, which indicated that three models possess good discriminative and generalization capability on unseen data, with XGBoost significantly outperforming the others ($P < 0.001$). The receiver operating characteristic curve (ROC) graphs were plotted using the "ggplot2" package in R (Fig. 2). The shapes of ROC graphs were similar among three datasets, indicating that the predictive ability of three models remained consistent across the training, validation, and test sets. All models had the highest AUC values on the training set, followed by the validation set, and the lowest on the test set, aligning with the general pattern of model generalization. LightGBM had the highest AUC values on the training and validation sets, while XGBoost had the highest AUC on the test set, with the smallest differences in AUC across the datasets. This suggests that LightGBM may be overfitting to the training data, while XGBoost is more robust and has stronger generalization ability. The recall/sensitivity value of XGBoost was 0.789 ± 0.026; LightGBM was 0.739 ± 0.030; CatBoost was 0.711 ± 0.030, suggesting that XGBoost was more sensitive to ED samples. The F1-Scores value of XGBoost was 0.695 ± 0.023, LightGBM was 0.681 ± 0.025; CatBoost was 0.681 ± 0.025. This indicated that the XGBoost model has better balanced precision, recall/sensitivity than the other two models ($P < 0.01$). The KAPPA value of XGBoost was 0.555 ± 0.032; LightGBM was 0.543 ± 0.033; CatBoost was 0.549 ± 0.034. All models had moderate consistency, with XGBoost significantly outperformed the other two models ($P < 0.01$).

Decision curve analysis (DCA) curves were plotted using the rmda package in R to evaluate the clinical utility of the three models (Fig. 3). As shown in Fig. 3, the standardized net benefit threshold range for applying the three models lied between 0 and 80%. Beyond a threshold of 80%, the benefits of employing the three models became less pronounced, even turning negative. The benefits brought by using the three models to guide clinical interventions exceed those from direct interventions across all threshold levels. The shapes of the DCA curves for the three models were similar, with the area under the curve of XGBoost slightly larger than that of the other two models. This suggested that choosing the XGBoost prediction model could offer the greatest benefit in most threshold scenarios.

The calibration curves were plotted to further evaluate the models' accuracy using the "rms" package in R (Fig. 4). As depicted in Fig. 4, the curves of the three models exhibited a generally monotonic increasing trend. When the predicted probability was less than 25%, the XGBoost curve almost overlapped with the ideal calibration line, indicating a higher prediction accuracy for non-ED samples. When the predicted probability was in 25-50% range, the LightGBM and CatBoost curve were closer to the ideal calibration line than XGBoost curve, which tended to overestimate the actual risk of ED occurrence, ,making its predictions more sensitive. When the predicted probability was above 50%, both the CatBoost and LightGBM curves were closer to the ideal calibration line, reflecting higher prediction accuracy for ED samples. Overall, the LightGBM curve performed the best, with the smallest Brier score loss (0.139) among the three models.

Considering these results collectively, we believed that XGBoost model is better than the other two models in terms of discriminative ability and clinical utility, with higher F1-score and Recall. The application scenario of the prediction model in this study focuses on the preliminary identification of ED. Although LightGBM model performed better in terms of accuracy, in the disease screening stage, correctly distinguishing patients and avoiding missed diagnoses are the primary objectives. Therefore, we concluded that the XGBoost model is the optimal ED prediction model in this study.

Feature importance and interpretability analysis

The 23 features of the XGBoost model were ranked according to their impact on the prediction results (Fig. 5). The top five most important features influencing the model's predictions, in order, are: age, diabetes, hemoglobin, PIR, and education level. Among these, Age has the greatest impact on the XGBoost algorithm, with a gain value of 0.55, significantly surpassing the importance of other features.

Post-hoc interpretations of the XGBoost model outputs were conducted using Shapley values (SHAP values) based on cooperative game theory, analyzing the marginal contribution of the features to the model output and ranking them accordingly. A SHAP summary plot was drawn (Fig. 6). In the figure, each point represents a specific single sample, with its position reflecting the impact of the corresponding feature on the predicted probability. A positive SHAP value indicates that the feature increases the predicted probability of having ED, while a negative value indicates a decrease. The larger the absolute value of the SHAP value, the more significant the impact of this feature on the predictive probability of ED. The color of the point indicates the corresponding feature value: lighter colors represent higher feature values. Continuous variables are displayed with gradient colors, indicating continuous changes from low to high values, while categorical variables have higher contrast colors that clearly distinguish different categories. Feature values are presented in Table 1.

As shown in Fig. 6, the data points for the features were distributed on both sides of the "0" value axis. This indicated that each feature had a bidirectional influence on the model's predicted probability of ED, depending on its value. The five features contributing the most to the model were age, PIR, education level, waist circumference

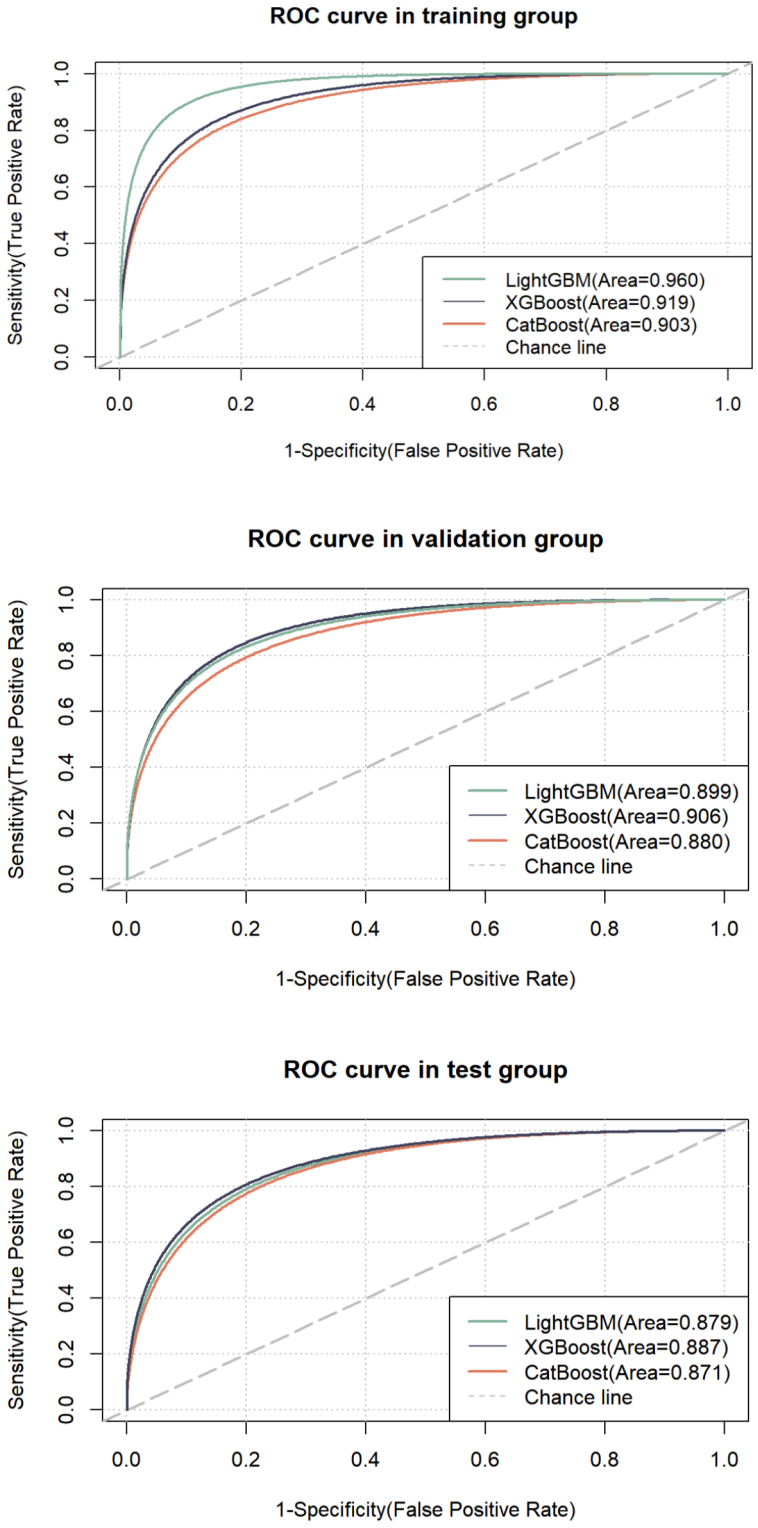

Fig. 2. ROC curve in training group, validation group and test group.

and shortness of breath on stairs/inclines. Among these, age and waist circumference are positively correlated with the predictive probability of ED; PIR is negatively correlated with the predictive probability of ED; respondents with shortness of breath on stairs or lower education levels are more likely to suffer from ED.

By analyzing the SHAP dependence plots of features (Supplementary Figure S2), it was found that among continuous variable features, all showed a nonlinear relationship with the predicted probability of ED. Among categorical variable features, those with a significant impact on the predicted probability of ED include diabetes,

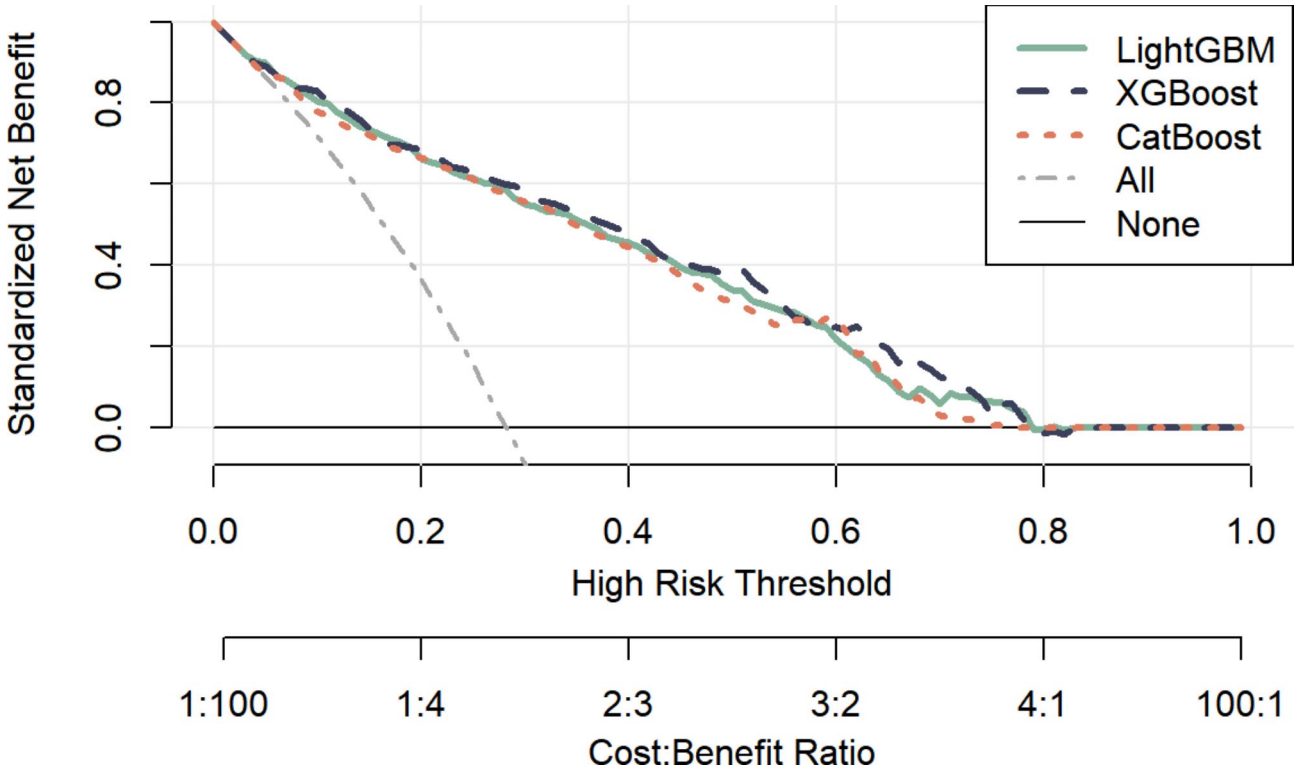

Fig. 3. Decision Curve Analysis (DCA) curves.

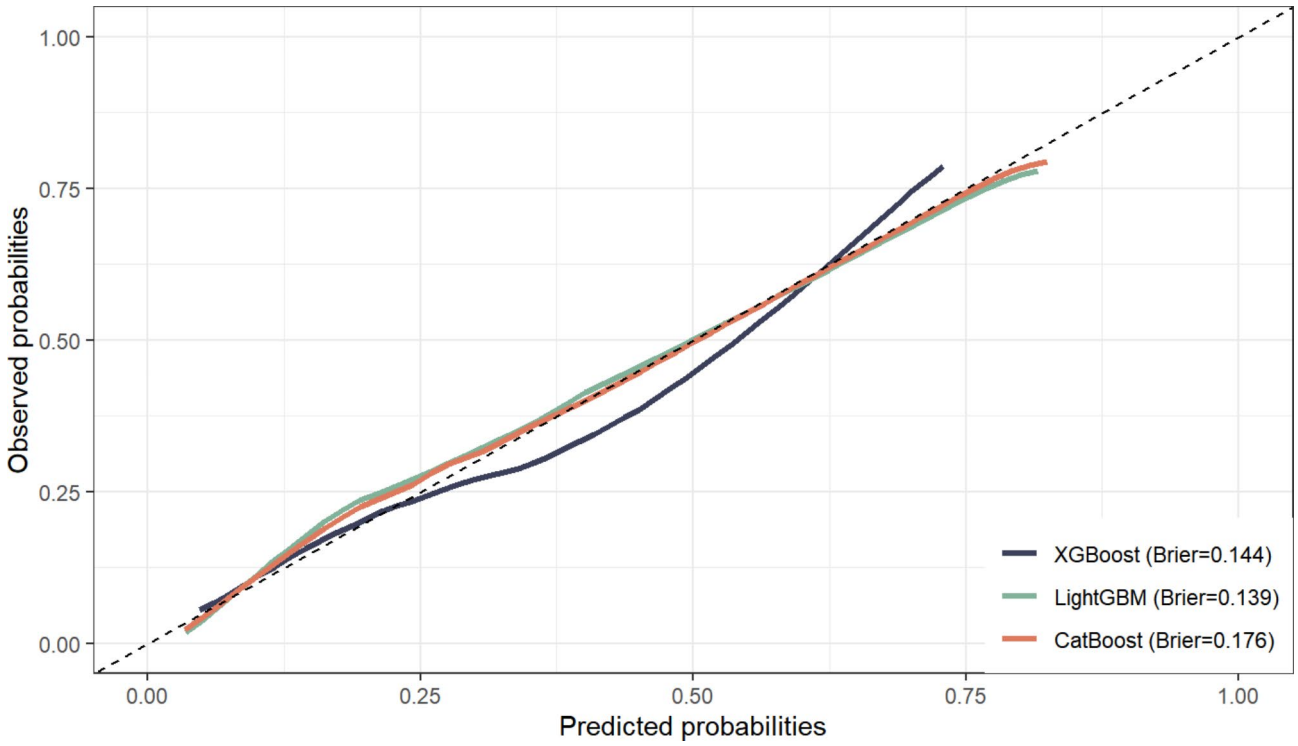

Fig. 4. Calibration curves.

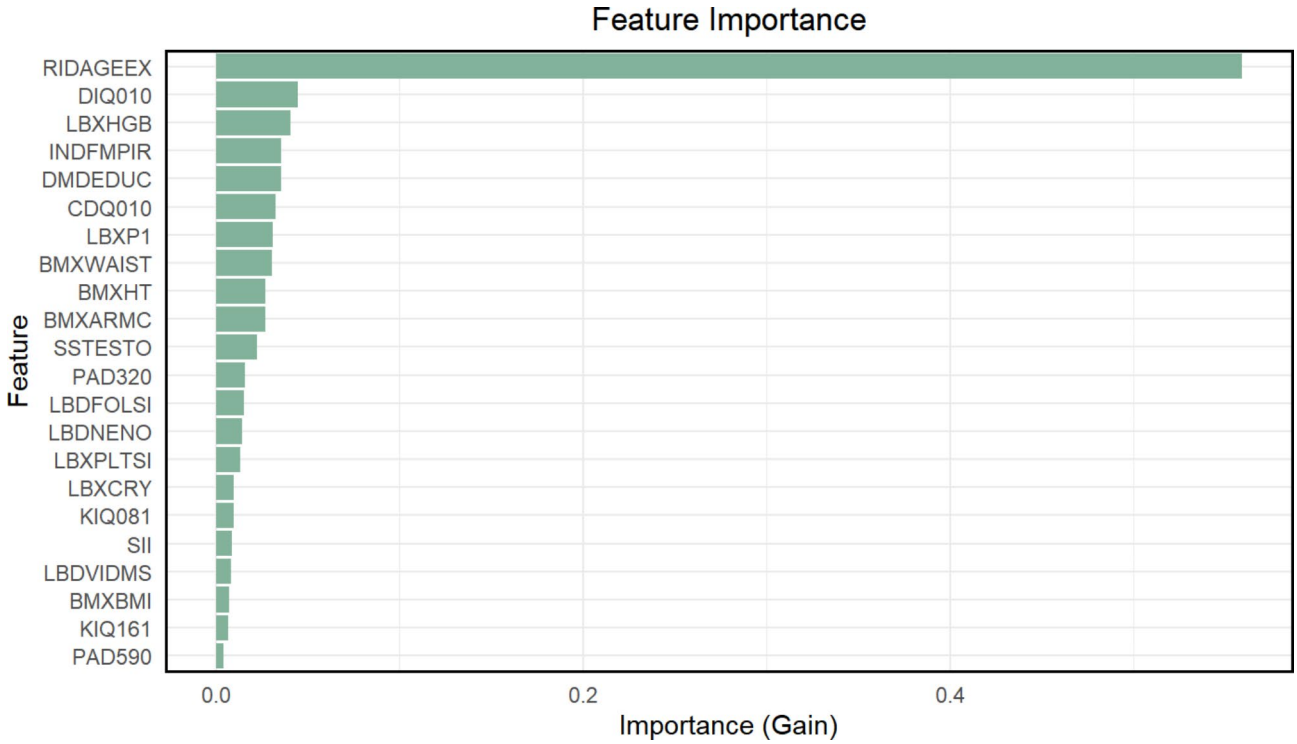

Fig. 5. Feature importance plot **NOTE**: RIDAGEEX: Age; DMDEDUC: Education level; INDFMPIR: Ratio of family income to poverty threshold; BMXBMI: Body Mass Index; BMXWAIST: Waist circumference; BMXARMC: Arm circumference; BMXHT: Height; SSTESTO: Testostosterone; PAD590: hours watch TV or videos past 30 days; PAD320: Moderate activity over past 30 days; DIQ010: Diabetes; LBXCRY: Beta-cryptoxanthin; LBDFOLSI: Folate; LBDVIDMS: Vitamin D; KIQ081: Usually have trouble trying to urinate; KIQ161: Age at first BPH diagnosis; LBXP1: PSA; LBDLYMNO: Lymphocyte number; LBXHGB: Hemoglobin; LBDNENO: Segmented neutrophils number; LBXPLTSI: Platelet count; SII: Systemic immune-inflammation index; CDQ010: Shortness of breath on stairs/inclines.

education level, shortness of breath on stairs/inclines, hours watch TV or videos past 30 days, moderate activity over past 30 days, and trouble trying to urinate.

Sensitivity analysis

This study explored the impact of various data preprocessing methods on the performance of the XGBoost model through multiple sensitivity analyses (Tables 2, 3, 4 and 5). The analysis results indicated that Logarithmic Transformation, feature selection methods, imputation techniques and SMOTE significantly affect model performance.

Discussion

This study, leveraging the multidimensional clinical features from NHANES, constructed an ED prediction model using the advanced GBDT algorithm, XGBoost. The model demonstrated excellent predictive performance, with high AUC value, sensitivity, and F1-Score, indicating that using this prediction model to assist in ED diagnosis could achieve a high rate of accuracy with low rates of missed diagnoses. Moreover, the model had strong clinical utility; firstly, its included features are common and easily obtainable, lowering the barrier to use. It could be applied in primary health care or self-assessment by patients, aiding in the early detection of ED, enhancing the awareness of ED among both patients and doctors, increasing the diagnosis rate of ED, and reducing the future risk of CVDs. Secondly, the model's DCA curve showed a net benefit above "all intervention" and "no intervention" strategies across the threshold range of 0–80% (cost-benefit ratio range of 1:100 to 4:1), meaning that aside from cases where patients strongly reject the potential adverse effects of misdiagnosis (such as treatment cost, time cost, side effects, and psychological anxiety), the use of this model in clinical decision-making is likely to bring higher benefits to both doctors and patients[28].

Several studies have emerged that use AI technology and machine learning techniques to assist in the diagnosis of ED[29–34]. In 2019, Chen et al[29]. trained an ED prediction model based on the Taiwan NHIRD database to assist clinical decisions. Their model, primarily featuring data related to ED comorbidities, included 19 features but did not incorporate demographic features and important laboratory features like PSA. In terms of predictive performance, their model reached a maximum AUC of 0.812, which is lower than the model in our study (0.887). In 2022, Zhang et al[35]. developed a model combining Support Vector Machine (SVM) and Recursive Feature Elimination (RFE) algorithms based on resting state functional magnetic resonance imaging (rs-fMRI)

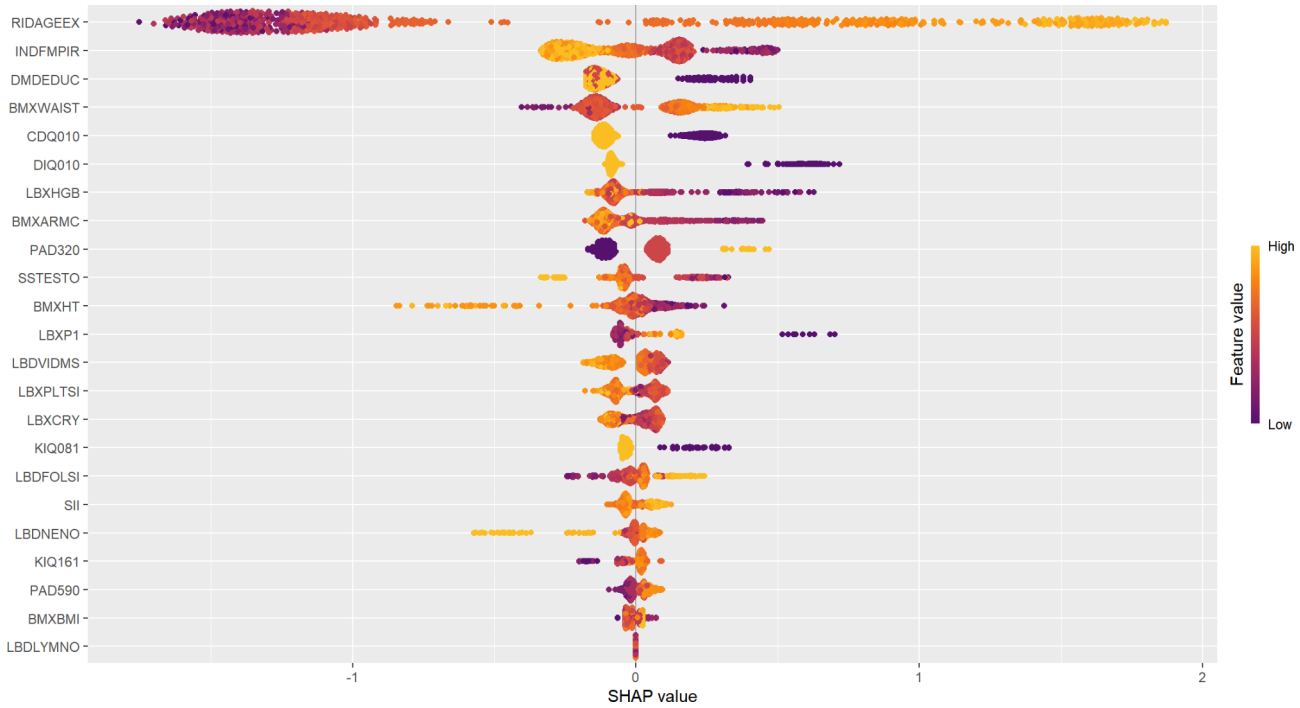

Fig. 6. SHAP summary plot **NOTE**: RIDAGEEX: Age; DMDEDUC: Education level; INDFMPIR: Ratio of family income to poverty threshold; BMXBMI: Body Mass Index; BMXWAIST: Waist circumference; BMXARMC: Arm circumference; BMXHT: Height; SSTESTO: Testostosterone; PAD590: hours watch TV or videos past 30 days; PAD320: Moderate activity over past 30 days; DIQ010: Diabetes; LBXCRY: Beta-cryptoxanthin; LBDFOLSI: Folate; LBDVIDMS: Vitamin D; KIQ081: Usually have trouble trying to urinate; KIQ161: Age at first BPH diagnosis; LBXP1: PSA; LBDLYMNO: Lymphocyte number; LBXHGB: Hemoglobin; LBDNENO: Segmented neutrophils number; LBXPLTSI: Platelet count; SII: Systemic immune-inflammation index; CDQ010: Shortness of breath on stairs/inclines.

Logarithmic Transformation	No. Features in Model	AUC	F1 _pos	F1- neg	Sensitivity	Specificity	Accuracy	KAPPA
Yes	23	0.887 ± 0.016	0.695 ± 0.023	0.857 ± 0.012	0.789 ± 0.026	0.811 ± 0.017	0.805 ± 0.015	0.555 ± 0.032
No	24	0.879 ± 0.017	0.680 ± 0.024	0.843 ± 0.013	0.794 ± 0.026	0.788 ± 0.018	0.789 ± 0.015	0.528 ± 0.032
***P*-Value**		< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001

Table 2. Sensitivity analysis result– logarithmic Transformation. The *P*-value indicates the significance level of intergroup differences after conducting T test /Wilcoxon test, with <0.05 indicating a significant difference.

Metrics	NOIMP vs. RF			NOIMP vs. RF + SMOTE			RF vs. RF + SMOTE			ANOVA *P*-value
	NOIMP	RF	*P*-value	NOIMP	RF + SMOTE	*P*-value	RF	RF + SMOTE	*P*-value	
AUC	0.884 ± 0.016	0.887 ± 0.016	< 0.001	0.884 ± 0.016	0.869 ± 0.018	< 0.001	0.887 ± 0.016	0.869 ± 0.018	< 0.001	< 0.001
F1-score_pos	0.686 ± 0.024	0.695 ± 0.023	< 0.001	0.686 ± 0.024	0.680 ± 0.024	< 0.001	0.695 ± 0.023	0.680 ± 0.024	< 0.001	< 0.001
F1-score_neg	0.853 ± 0.012	0.857 ± 0.012	< 0.001	0.853 ± 0.012	0.853 ± 0.012	0.7559	0.857 ± 0.012	0.853 ± 0.012	< 0.001	< 0.001
Sensitivity	0.775 ± 0.027	0.789 ± 0.026	< 0.001	0.775 ± 0.027	0.762 ± 0.028	< 0.001	0.789 ± 0.026	0.762 ± 0.028	< 0.001	< 0.001
Specificity	0.809 ± 0.017	0.811 ± 0.017	< 0.05	0.809 ± 0.017	0.813 ± 0.017	< 0.001	0.811 ± 0.017	0.813 ± 0.017	0.15	< 0.001
Accuracy	0.800 ± 0.015	0.805 ± 0.015	< 0.001	0.800 ± 0.015	0.798 ± 0.015	< 0.001	0.805 ± 0.015	0.798 ± 0.015	< 0.001	< 0.001
KAPPA	0.541 ± 0.032	0.555 ± 0.032	< 0.001	0.541 ± 0.032	0.535 ± 0.033	< 0.001	0.555 ± 0.032	0.535 ± 0.033	< 0.001	< 0.001

Table 3. Sensitivity analysis result — SMOTE. The *P*-value indicates the significance level of intergroup differences after conducting Tukey HSD /Games-Howell/Dunn test, with < 0.05 indicating a significant difference. RF: Using random forest method for imputing missing values; NOIMP: Not imputing missing values; RF + SMOTE: Using random forest method for imputing missing values, and employ the SMOTE + ENN to oversample the model training dataset in each iteration before training commenced.

Metrics	NOIMP vs. PMM			NOIMP vs. RF			PMM vs. RF			ANOVA *P*-value
	NOIMP	PMM	*P*-value	NOIMP	RF	*P*-value	PMM	RF	*P*-value	
AUC	0.884 ± 0.016	0.869 ± 0.018	< 0.001	0.884 ± 0.016	0.887 ± 0.016	< 0.001	0.869 ± 0.018	0.887 ± 0.016	< 0.001	< 0.001
F1-score_pos	0.686 ± 0.024	0.675 ± 0.025	< 0.001	0.686 ± 0.024	0.695 ± 0.023	< 0.001	0.675 ± 0.025	0.695 ± 0.023	< 0.001	< 0.001
F1-score_neg	0.853 ± 0.012	0.847 ± 0.012	< 0.001	0.853 ± 0.012	0.857 ± 0.012	< 0.001	0.847 ± 0.012	0.857 ± 0.012	< 0.001	< 0.001
Sensitivity	0.775 ± 0.027	0.765 ± 0.029	< 0.001	0.775 ± 0.027	0.789 ± 0.026	< 0.001	0.765 ± 0.029	0.789 ± 0.026	< 0.001	< 0.001
Specificity	0.809 ± 0.017	0.803 ± 0.017	< 0.001	0.809 ± 0.017	0.811 ± 0.017	< 0.05	0.803 ± 0.017	0.811 ± 0.017	< 0.001	< 0.001
Accuracy	0.800 ± 0.015	0.793 ± 0.015	< 0.001	0.800 ± 0.015	0.805 ± 0.015	< 0.001	0.793 ± 0.015	0.805 ± 0.015	< 0.001	< 0.001
KAPPA	0.541 ± 0.032	0.526 ± 0.034	< 0.001	0.541 ± 0.032	0.555 ± 0.032	< 0.001	0.526 ± 0.034	0.555 ± 0.032	< 0.001	< 0.001

Table 4. Sensitivity analysis result — imputation techniques. The *P*-value indicates the significance level of intergroup differences after conducting Tukey HSD /Games-Howell/Dunn test, with < 0.05 indicating a significant difference. RF: Using random forest method for imputing missing values; PMM: Using the Predictive Mean Matching (PMM) for numerical variables and mode imputation for categorical variables to impute missing values; NOIMP: Not imputing missing values.

Metrics	RF25 vs. RF20			RF25 vs. RF15			RF20 vs. RF15			ANOVA *P*-value
	RF25	RF20	*P*-value	RF25	RF15	*P*-value	RF20	RF15	*P*-value	
AUC	0.887 ± 0.016	0.868 ± 0.017	< 0.001	0.887 ± 0.016	0.871 ± 0.017	< 0.001	0.868 ± 0.017	0.871 ± 0.017	< 0.001	< 0.001
F1-score_pos	0.695 ± 0.023	0.676 ± 0.024	< 0.001	0.695 ± 0.023	0.665 ± 0.024	< 0.001	0.676 ± 0.024	0.665 ± 0.024	< 0.001	< 0.001
F1-score_neg	0.857 ± 0.012	0.847 ± 0.012	< 0.001	0.857 ± 0.012	0.838 ± 0.013	< 0.001	0.847 ± 0.012	0.838 ± 0.013	< 0.001	< 0.001
Sensitivity	0.789 ± 0.026	0.770 ± 0.027	< 0.001	0.789 ± 0.026	0.770 ± 0.027	< 0.001	0.770 ± 0.027	0.770 ± 0.027	0.99	< 0.001
Specificity	0.811 ± 0.017	0.801 ± 0.017	< 0.001	0.811 ± 0.017	0.786 ± 0.018	< 0.001	0.801 ± 0.017	0.786 ± 0.018	< 0.001	< 0.001
Accuracy	0.805 ± 0.015	0.792 ± 0.015	< 0.001	0.805 ± 0.015	0.782 ± 0.015	< 0.001	0.792 ± 0.015	0.782 ± 0.015	< 0.001	< 0.001
KAPPA	0.555 ± 0.032	0.526 ± 0.033	< 0.001	0.555 ± 0.032	0.508 ± 0.033	< 0.001	0.526 ± 0.033	0.508 ± 0.033	< 0.001	< 0.001

Table 5. Sensitivity analysis result — feature selection methods. The *P*-value indicates the significance level of intergroup differences after conducting Tukey HSD /Games-Howell/Dunn test, with < 0.05 indicating a significant difference. RF25: Select the top 25 most important features from each model and take the intersection; RF20: Select the top 20 most important features from each model and take the intersection; RF15: Select the top 15 most important features from each model and take the intersection.

data features of Psychogenic erectile dysfunction (pED) patients under 40 and a corresponding healthy control group (HCS). In terms of performance metrics, their model demonstrated superior performance compared to ours (AUC of 0.9301, sensitivity of 100%, and specificity of 0.88). However, their study used a smaller and unbalanced dataset (pED: 48, HCS: 39) and employed leave-one-out cross-validation (LOOCV), implying the training dataset is approximately equal to the entire dataset, which might lead to overfitting to the training data. In addition, the model's performance in the context of an imbalanced dataset (more positive than negative subjects) did not discuss the F1-Score, overlooking the balance between precision and recall, which diminishing the credibility of the model. Our model's F1-Score was 0.695, which reflected that the model balanced the rates of missed and false diagnoses relatively well, thereby improving credibility. In recent years, AI chatbots like ChatGPT have become an important source of medical information for the public. Studies have found that the quality of ED-related health information provided by AI chatbots is generally high. However, there are currently no reports of using AI chatbots to assist in ED diagnosis[36].

This study employed various data preprocessing methods to enhance model performance. Logarithmic transformation before random forest (RF) imputation significantly improved the performance of the XGBoost model across most metrics ($P < 0.001$), indicating that logarithmic transformation could balance data distribution and stabilize data variance, allowing the XGBoost model to more accurately identify the patterns and structure in the dataset. Although sensitivity slightly decreased compared to the model without logarithmic transformation, the overall benefits of logarithmic transformation were clearly higher. Missing Data can affect model performance[37]; the NHANES analysis guidelines recommend appropriately imputing missing data to reduce bias[25]. It is also noteworthy that GBDT algorithms like XGBoost have built-in capabilities for handling missing data. Therefore, we conducted sensitivity analysis to clarify the impact of different imputation techniques on model performance. The analysis showed that RF imputation method was the most effective imputation technique in this study, significantly outperforming both non-imputation and PMM plus mode imputation method ($P < 0.001$). This suggested that RF method helped retain the inherent structure and relationships within the data, making it suitable for NHANES nonlinear data. The model trained on the dataset imputed by the PMM plus mode method performed worse than the one trained with RF method and even worse than the one trained on the non-imputed dataset. This suggested that the PMM, which tends to simplify the data structure, may not be suitable for the dataset in this study. It also highlighted the strong built-in capabilities of XGBoost for handling missing data, which warrants further discussion. The class imbalance is a very prevalent

problem in machine learning models, and the SMOTE + ENN technique is a common solution for imbalance issues[38,39]. However, sensitivity analysis results showed that applying SMOTE + ENN technique to RF-imputed data resulted in decreased F1-scores and sensitivity, while specificity increased slightly but not significantly ($P = 0.15$). This suggests that the SMOTE + ENN may have introduced noise or caused overfitting, making it unsuitable for this study. Feature selection can influence the model performance, and should strike a balance between retaining important features and avoiding overfitting. Sensitivity analysis of feature selection methods showed that the number of selected features was directly proportional to the XGBoost model's performance. The feature selection method that took the top 25 features from each model and used their intersection provided the best results for the XGBoost model, with no clear tendency towards overfitting, making it an effective feature selection method. In summary, the data preprocessing methods employed effectively enhance the model's performance and improve the model's interpretability, playing an important role in our study.

The feature set of this ED prediction model was comprehensive. We further analyzed the interpretability of the model (Fig. 6, Supplementary Figure S2) and found each feature has significant clinical relevance in the pathophysiological mechanism of ED. ED is closely related to CVD, shortness of breath on stairs/inclines is one of the main manifestations of CVD. The results of the SHAP dependence plot (Supplementary Figure S2 D) in this study indicated that Shortness of breath on stairs/inclines significantly increases the predicted probability of ED. Oxidative stress may be the common pathological mechanism between CVD and ED, which plays an important role in the pathogenesis of vascular ED[8,40]. Sexual stimulation leads to the release of nitric oxide (NO) and other endothelial factors into the blood. These endothelial factors determine the relaxation of the arterial smooth muscle supplying blood to the corpus cavernosum of the penis, increasing penile blood flow while also blocking venous outflow. This significantly increases the blood volume and pressure within the corpora cavernosum, leading to an erection. During periods of oxidative stress, an excess production of reactive oxygen species (ROS) leads to a significant reduction in vascular NO, thereby affecting the endothelium-dependent vasodilation process and causing ED. Studies have shown that conditions such as hypertension, obesity, hyperglycemia, and hypercholesterolemia can lead to a rapid increase in ROS, thus triggering oxidative stress in the vascular system[41]. This suggests that these diseases are not only risk factors for cardiovascular diseases but also for ED. Patients with diabetes[42,43], hypertension[43,44], and hypercholesterolemia[45]have a higher incidence rate of ED, and their condition of ED tends to be more severe compared to those without cardiovascular risk factors. According to the interpretability analysis (Supplementary Figure S2 E), diabetes significantly increased the predicted probability of ED, consistent with the findings of real-data research studies. Inflammation also plays an important role in endothelial dysfunction. The elevation of SII, as a combined indicator of peripheral blood lymphocytes, neutrophils, and platelets, indicates aggravated inflammation and increased endothelial damage[46]. The studies[47,48]published in 2024 found a nonlinear correlation between SII and ED, SII > 485.530 was correlated with an increased risk of ED, consistent with the findings of our study (Supplementary Figure S2 T). Beta-cryptoxanthin and vitamin D can regulate vascular endothelial function. Beta-cryptoxanthin, a type of carotenoid, has strong antioxidant properties. Lower levels of serum beta-cryptoxanthin have been linked to ED[49]. Vitamin D is an important predictor of CVD, and several studies have shown a link between decreased vitamin D levels and ED[50]. In this study, the SHAP dependence plots (Supplementary Figure S2 K, M) showed that the levels of beta-cryptoxanthin and vitamin D were significantly associated with the predicted probability of ED. When beta-cryptoxanthin > 8 μg/dL and vitamin D > 70 nmol/L, their contribution to the predicted probability of ED was mainly negative, providing a protective effect. These findings align with the definition of adequate vitamin D levels (25-hydroxyvitamin D level reach 75–250 nmol/L) in guideline[51], suggesting that appropriate supplementation of vitamins such as vitamin D and beta-cryptoxanthin may help protect vascular endothelial function and prevent the occurrence of ED.

Androgens, especially testosterone, are essential for the pathophysiological mechanisms of ED[52,53]. Studies have demonstrated that ED is maintained by the levels of plasma testosterone, which can promote the synthesis of NO and NOS, alter the structure of smooth muscle, connective tissue, and nerve fibers in the corpus cavernosum, thereby affecting ED[52–54]. When testosterone levels decrease, men may not only experience ED but also a range of symptoms including reduced muscle mass, obesity, and cognitive decline, collectively referred to as Late-Onset Hypogonadism (LOH)[55]. According to the 2018 guidelines by The American Urological Association (AUA)[56]and the 2019 guidelines by the European Association of Urology (EAU)[57], the diagnostic threshold for low testosterone related to ED is set at 244.14 ng/dL. The results of the SHAP dependence plot (Supplementary Figure S2 U) in this study indicated that the contribution to the predicted probability of ED was mainly positive when serum testosterone levels were below 2.5 ng/ml (= 250 ng/dl), which is consistent with the guideline recommendations. Regular physical activities can help maintain muscle mass. The results of the SHAP dependence plot (Supplementary Figure S2 P) in this study showed that a decrease in moderate activity significantly increased the predicted probability of ED, which may be related to the reduced muscle mass caused by long-term lack of exercise. A decrease in arm circumference is also closely associated with reduced muscle mass, being one of the significant manifestations of LOH and sarcopenia[58,59]. Patients with sarcopenia may experience ED, with a smaller arm circumference correlating with more severe ED[60]. The underlying mechanism may be related to the reduction in testosterone levels. A 2019 study indicated that patients without ED had an average arm circumference greater than 32 cm, while those with mild and severe ED had an average arm circumference less than 32 cm[61]. The interpretability analysis suggested a non-linear negative correlation between arm circumference and the predictive probability of ED, where an arm circumference smaller than 32 cm could significantly increase the predictive probability of ED (Supplementary Figure S2 A), aligning with the findings from the 2019 study. Prostate cancer is closely associated with ED, with over 80% of prostate cancer patients experiencing erectile dysfunction[62]. This could be related to cancer-induced malnutrition and somatization disorders, as well as the treatment regimens for cancer. Androgen deprivation therapy (ADT) can lower androgen levels in cancer patients, leading to ED symptoms during treatment. PSA is a screening

marker for the diagnosis of prostate cancer; higher PSA levels increase the likelihood of prostate cancer, thereby increasing the incidence rate of ED. Our study showed that an increase in PSA was positively correlated with an increased predictive probability of ED, consistent with the conclusions mentioned above (Supplementary Figure S2 S).

Benign Prostatic Hyperplasia/Lower Urinary Tract Symptoms (BPH/LUTS) and ED may share similar pathophysiological mechanisms[63]. Oxidative stress can lead to endothelial dysfunction, resulting in decreased nitric oxide (NO) and vasodilator factors. This may cause vasoconstriction, which then reduces blood flow to pelvic organs and nerves, leading to tissue hypoxia. Hypoxia can alter the contractility of the penile corpora cavernosum and lower urinary tract smooth muscle, as well as affect the neural regulation of lower urinary tract and sexual function. Phosphodiesterase type 5 (PDE5) inhibitors can promote the relaxation of blood vessels and smooth muscle, improving blood flow and oxygenation to lower urinary tract tissues while alleviating LUTS and ED[64–66]. Research has confirmed a significant correlation between ED and BPH/LUTS[53,67]. The interpretability analysis regarding urinate trouble also confirmed the correlation between the two conditions, with BPH/LUTS significantly increasing the predictive probability of ED (Supplementary Figure S2 V). Moreover, the analysis indicated that hemoglobin is negatively correlated with the predictive probability of ED, indirectly confirming the mechanism by which hypoxemia contributes to the incidence of ED (Supplementary Figure S2 L).

Age is the demographic feature most strongly associated with ED and makes the largest contribution to the prediction model. As age increases, endothelial function gradually deteriorates, and testosterone levels gradually decrease, which in various ways increase the prevalence of ED[68]. The interpretability analysis revealed that the relationship between age and the predictive probability of ED was nonlinear, with a significant positive correlation after the age of 40, and the predictive probability of ED reaching 50% around the age of 50 (Supplementary Figure S2 B). This matched the findings of the Massachusetts Male Aging Study (MMAS) on the prevalence of ED and a 2014 BMJ study on the age distribution of ED[69]. Similar to age, obesity also plays a role in the pathophysiological mechanisms of ED[70,71]. Existing research has confirmed that waist circumference, compared to BMI, is a better indicator of central obesity and visceral fat level, with higher waist circumference indicating a higher risk of ED[72]. According to the SHAP dependence plot (Supplementary Figure S2 W), a waist circumference > 100 cm significantly increases the predictive probability of ED, consistent with current research[73,74].

In addition, among demographic characteristics, the education level and PIR are important components of Socioeconomic Status (SES), contributing significantly to this ED prediction model. PIR showed a nonlinear negative correlation with the predictive probability of ED, while a lower education level (Less Than High School) can significantly increase the predictive probability of ED (Supplementary Figure S2 F, R). This indicated that individuals with lower SES have a higher predictive probability of ED. Previous research has confirmed SES as a risk factor for ED, though the specific mechanisms remain unclear[75]. This may be related to potential psychological stress and lower rates of seeking medical advice associated with low SES[76–78]. In summary, the features included in this ED prediction model are all clinically meaningful, and the cutoff values based on SHAP dependence plots are consistent with authoritative clinical research and guidelines, indicating the model's strong interpretability.

This study has certain limitations. The model has performance limitations: although the XGBoost prediction model can adequately meet the needs for preliminary screening of ED patient, its accuracy (Brier loss score = 0.144, PPV = 0.622 ± 0.029) still has room for improvement and is currently insufficient for precise prediction of the probability of ED. Although we made effort to avoid data leakage during the methods' design, including but not limited to strictly separating the train set (80%) and test set (20%), employing consistent imputation techniques across these sets, and separating the target variable during RF imputation, the feature selection performed on the entire train set (80%) before conducting 5-fold cross-validation could result in the model's performance on the validation set being higher than actual. The current method of feature selection, while ensuring comparability among models in sensitivity analysis, may lead to the omission of some important features (such as hypertension, hyperlipidemia) by using the intersection method. Due to budget constraints, we did not use datasets other than NHANES for external validation of the model, which limited our ability to assess the model's generalizability across different population structures. Doctors are the primary users of this model; however, they currently possess a limited understanding of ED and often neglect patients' sexual health needs. Consequently, high-risk ED patients, particularly older individuals, frequently do not receive effective evaluation and treatment[14]. This could potentially restrict the willingness to use and the applicability of the model. Clinically, ED diagnosis is often made using detailed criteria such as the International Index of Erectile Function (IIEF-5)[30], which provides a comprehensive ED assessment and grades its severity levels. In contrast, the NHANES dataset offers a more simplified ED assessment, limiting the model's application in clinical decision. The NHANES dataset from 2001 to 2004 was used for model training, suggesting the model obtained may have time limitations. Taking diabetes as an example, it plays a significant role in this model, and the risk of ED is higher in samples with diabetes. Epidemiological studies have shown a significant increase in the prevalence of diabetes in the U.S. population over the last 30 years. Additionally, other research indicates that recently marketed diabetes medications, such as exenatide, are effective in managing ED symptoms[79–81]. Changes in patient demographics and advancements in treatment technologies over time may have altered the importance of features, such as diabetes, in the ED prediction model, affecting the accuracy of the model's predictions. The incidence rate of ED is increasing among young people, with as high as 30% in young populations[82]. When ED occurs in young people, the risk of future cardiac events increases significantly. However, due to the small number of ED samples in this study, we were unable to stratify data by age (20–40, 40–60, 60 and above) to train the model, preventing an analysis of factors related to ED incidence in different age groups. Future studies could use databases from different countries and regions for external validation to enhance the model's applicability and generalizability; enhance education for doctors to emphasize the significant role of ED in cardiovascular, endocrine, and other related diseases, in order to increase the usage rate of the model and further improve the diagnosis and treatment rates

for ED patients; refine ED assessment criteria to further develop multicategory models for the severity of ED, deeply involving clinical treatment decisions; update the feature data in the model before clinical application to ensure its applicability and accuracy in specific populations; and create ED prediction models for young people to explore risk factors for ED incidence in young people and reduce the incidence rate of cardiac events.

Conclusion

Compared to current related research models, our prediction model has the following advantages: firstly, it has stronger predictive capabilities (has higher AUC, higher F1-Score); secondly, it offers greater interpretability (contains multiple features, all involved in the ED pathophysiological mechanism); and lastly, it has stronger generalizability (contains features easily accessible). Our model can broaden the range of diagnostic methods for ED, enhance the diagnosis rate of ED, and support doctors in implementing early interventions for ED patients, thereby significantly improving patients' life quality.

Data availability

The survey data are publicly available on the Internet for data users and researchers throughout the world (www.cdc.gov/nchs/nhanes/).The R implementation of the study is available on GitHub (https://github.com/Aquapop/ML_ED).

Received: 2 April 2024; Accepted: 4 November 2024
Published online: 08 November 2024

References

1. Ayta, I. A., McKinlay, J. B. & Krane, R. J. The likely worldwide increase in erectile dysfunction between 1995 and 2025 and some possible policy consequences. *BJU Int.* **84**, 50–56. https://doi.org/10.1046/j.1464-410x.1999.00142.x (1999).
2. Burnett, A. L. et al. Erectile Dysfunction: AUA Guideline. *J. Urol.* **200**, 633–641. https://doi.org/10.1016/j.juro.2018.05.004 (2018).
3. McKinlay, J. B. The worldwide prevalence and epidemiology of erectile dysfunction. *Int. J. Impot. Res.* **12** (Suppl 4), S6–s11. https://doi.org/10.1038/sj.ijir.3900567 (2000).
4. Sugimori, H. et al. Relationships between erectile dysfunction, depression, and anxiety in Japanese subjects. *J. Sex. Med.* **2**, 390–396. https://doi.org/10.1111/j.1743-6109.2005.20354.x (2005).
5. Xiao, Y. et al. Factors associated with anxiety and depression in patients with erectile dysfunction: a cross-sectional study. *BMC Psychol.* **11**, 36. https://doi.org/10.1186/s40359-023-01074-w (2023).
6. Muneer, A., Kalsi, J., Nazareth, I. & Arya, M. Erectile dysfunction. *BMJ (Clinical Res. ed)*. **348**, g129. https://doi.org/10.1136/bmj.g129 (2014).
7. Barrett-Connor, E. & Heart Disease Risk Factors Predict Erectile Dysfunction 25 Years Later (The Rancho Bernardo Study). *Am. J. Cardiol.* **96**, 3–7 https://doi.org:https://doi.org/10.1016/j.amjcard.2005.07.012 (2005).
8. Gandaglia, G. et al. A systematic review of the Association between Erectile Dysfunction and Cardiovascular Disease. *Eur. Urol.* **65**, 968–978. https://doi.org/10.1016/j.eururo.2013.08.023 (2014). https://doi.org
9. Keenan, H. A. Do Erectile Dysfunction and Cardiovascular Disease have the same mechanism? *Eur. Urol.* **65**, 979–980. https://doi.org/10.1016/j.eururo.2013.11.013 (2014).
10. Joseph, P. et al. Reducing the Global Burden of Cardiovascular Disease, Part 1: the epidemiology and risk factors. *Circul. Res.* **121**, 677–694. https://doi.org/10.1161/circresaha.117.308903 (2017).
11. Gareri, P., Castagna, A., Francomano, D., Cerminara, G. & De Fazio, P. Erectile dysfunction in the elderly: an old widespread issue with novel treatment perspectives. *Int. J. Endocrinol,* 2014, 878670, 15 pages, 2014. https://doi.org/10.1155/2014/878670.
12. Lin, H., Zhao, L., Wu, H., Cao, M. & Jiang, H. Sexual life and medication taking behaviours in young men: an online survey of 92 620 respondents in China. *Int. J. Clin. Pract.* **74**, e13417 (2020).
13. Klaeson, K., Hovlin, L., Guvå, H. & Kjellsdotter, A. Sexual health in primary health care - a qualitative study of nurses' experiences. *J. Clin. Nurs.* **26**, 1545–1554. https://doi.org/10.1111/jocn.13454 (2017).
14. Rusiecki, L. et al. Sexual health in Polish elderly men with coronary artery disease: importance, expectations, and reality. *Asian J. Androl.* **22**, 526–531. https://doi.org/10.4103/aja.aja_131_19 (2020).
15. Kałka, D. et al. Should information about sexual health be included in education directed toward men with cardiovascular diseases? The aging male. *Official J. Int. Soc. Study Aging Male*. **21**, 243–250. https://doi.org/10.1080/13685538.2018.1439911 (2018).
16. Gao, Q. Q., Chen, J. H., Chen, Y., Song, T. & Dai, Y. T. Dynamic infusion cavernosometry and cavernosography for classifying venous erectile dysfunction and its significance for individual treatment. *Chin. Med. J.* **132**, 405–410. https://doi.org/10.1097/cm9.0000000000000099 (2019).
17. Jannini, E. A., Granata, A. M., Hatzimouratidis, K. & Goldstein, I. Use and abuse of Rigiscan in the diagnosis of erectile dysfunction. *J. Sex. Med.* **6**, 1820–1829. https://doi.org/10.1111/j.1743-6109.2009.01343.x (2009).
18. Oh, J. H. et al. Computational methods using genome-wide association studies to predict radiotherapy complications and to identify correlative molecular processes. *Sci. Rep.* **7**, 43381. https://doi.org/10.1038/srep43381 (2017).
19. Tsai, S. F., Yang, C. T., Liu, W. J. & Lee, C. L. Development and validation of an insulin resistance model for a population without diabetes mellitus and its clinical implication: a prospective cohort study. *EClinicalMedicine*. **58**, 101934. https://doi.org/10.1016/j.eclinm.2023.101934 (2023).
20. Zhang, C. et al. *J. Urol.* **199**, e935–e935 https://doi.org/10.1016/j.juro.2018.02.2248 (2018).
21. Shehab, M. et al. Machine learning in medical applications: a review of state-of-the-art methods. *Comput. Biol. Med.* **145**, 105458. https://doi.org/10.1016/j.compbiomed.2022.105458 (2022).
22. Collins, G. S., Reitsma, J. B., Altman, D. G. & Moons, K. G. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ (clinical research ed.) 350, g7594 (2015). https://doi.org/10.1136/bmj.g7594
23. O'Donnell, A. B., Araujo, A. B., Goldstein, I. & McKinlay, J. B. The validity of a single-question self-report of erectile dysfunction. Results from the Massachusetts Male Aging Study. *J. Gen. Intern. Med.* **20**, 515–519. https://doi.org/10.1111/j.1525-1497.2005.0076.x (2005).
24. Song, Y. et al. Systemic immune-inflammation index is associated with hepatic steatosis: evidence from NHANES 2015–2018. *Front. Immunol.* **13**, 1058779. https://doi.org/10.3389/fimmu.2022.1058779 (2022).
25. Johnson, C. L. et al. National health and nutrition examination survey. Analytic guidelines, 1999–2010. (2013).
26. Merianos, A. L. et al. Distinguishing Exposure to Secondhand and Thirdhand Tobacco Smoke among U.S. Children Using Machine Learning: NHANES 2013–2016. *Environ. Sci. Technol.* **57**, 2042–2053 https://doi.org/10.1021/acs.est.2c08121 (2023).
27. Zhang, Y. P. et al. Artificial intelligence-driven radiomics study in cancer: the role of feature engineering and modeling. *Military Med. Res.* **10**, 22. https://doi.org/10.1186/s40779-023-00458-8 (2023).

28. Vickers, A. J., van Calster, B. & Steyerberg, E. W. A simple, step-by-step guide to interpreting decision curve analysis. *Diagn. Prognostic Res.* **3**, 18. https://doi.org/10.1186/s41512-019-0064-7 (2019).
29. Chen, Y. F. et al. Design of a clinical decision support system for predicting erectile dysfunction in men using NHIRD dataset. *IEEE J. Biomedical Health Inf.* **23**, 2127–2137 (2018).
30. Glavaš, S. et al. Erectile function in cardiovascular patients: its significance and a quick assessment using a visual-scale questionnaire. *Acta Cardiol.* **70**, 712–719. https://doi.org/10.1080/AC.70.6.3120185 (2015).
31. Gorek, M., Stief, C. G., Hartung, C. & Jonas, U. Computer-assisted interpretation of electromyograms of corpora cavernosa using fuzzy logic. *World J. Urol.* **15**, 65–70. https://doi.org/10.1007/bf01275159 (1997).
32. Jang, I. et al. LC-MS/MS Software for Screening unknown erectile dysfunction drugs and analogues: Artificial neural network classification, peak-count scoring, simple similarity search, and hybrid similarity search algorithms. *Anal. Chem.* **91**, 9119–9128. https://doi.org/10.1021/acs.analchem.9b01643 (2019).
33. Kellner, B., Stief, C. G., Hinrichs, H. & Hartung, C. Computerized classification of corpus cavernosum electromyogram signals by the use of discriminant analysis and artificial neural networks to support diagnosis of erectile dysfunction. *Urol. Res.* **28**, 6–13. https://doi.org/10.1007/s002400050002 (2000).
34. Li, L. et al. Abnormal brain structure as a potential biomarker for venous erectile dysfunction: evidence from multimodal MRI and machine learning. *Eur. Radiol.* **28**, 3789–3800. https://doi.org/10.1007/s00330-018-5365-7 (2018).
35. Zhang, X. et al. Identifying neuroimaging biomarkers for psychogenic erectile dysfunction by fusing multi-level brain information: a resting-state functional magnetic resonance imaging study. *Andrology*. **10**, 1398–1410. https://doi.org/10.1111/andr.13238 (2022).
36. Pan, A., Musheyev, D., Loeb, S. & Kabarriti, A. E. Quality of erectile dysfunction information from ChatGPT and other artificial intelligence chatbots. *BJU Int.* **133**, 152–154. https://doi.org/10.1111/bju.16209 (2024).
37. de Goeij, M. C. M. et al. Multiple imputation: dealing with missing data. *Nephrol. Dialysis Transplantation*. **28**, 2415–2420. https://doi.org/10.1093/ndt/gft221 (2013).
38. Branco, P., Torgo, L. & Ribeiro, R. P. A survey of predictive modeling on imbalanced domains. *ACM Comput. Surv. (CSUR)*. **49**, 1–50 (2016).
39. Batista, G. E., Prati, R. C. & Monard, M. C. A study of the behavior of several methods for balancing machine learning training data. *ACM SIGKDD Explorations Newsl.* **6**, 20–29 (2004).
40. Agarwal, A., Nandipati, K. C., Sharma, R. K., Zippe, C. D. & Raina, R. Role of oxidative stress in the pathophysiological mechanism of erectile dysfunction. *J. Androl.* **27**, 335–347. https://doi.org/10.2164/jandrol.05136 (2006).
41. Förstermann, U. Oxidative stress in vascular disease: causes, defense mechanisms and potential therapies. *Nat. Clin. Pract. Cardiovasc. Med.* **5**, 338–349. https://doi.org/10.1038/ncpcardio1211 (2008).
42. Rosen, R. C. et al. The multinational men's attitudes to life events and sexuality (MALES) study: I. Prevalence of erectile dysfunction and related health concerns in the general population. *Curr. Med. Res. Opin.* **20**, 607–617. https://doi.org/10.1185/030079904125003467 (2004).
43. Kouidrat, Y. et al. High prevalence of erectile dysfunction in diabetes: a systematic review and meta-analysis of 145 studies. *Diabet. Medicine: J. Br. Diabet. Association*. **34**, 1185–1192. https://doi.org/10.1111/dme.13403 (2017).
44. Burchardt, M. et al. Hypertension is associated with severe erectile dysfunction. *J. Urol.* **164**, 1188–1191 (2000).
45. Seftel, A. D., Sun, P. & Swindle, R. The prevalence of hypertension, hyperlipidemia, diabetes mellitus and depression in men with erectile dysfunction. *J. Urol.* **171**, 2341–2345. https://doi.org/10.1097/01.ju.0000125198.32936.38 (2004).
46. Wang, L., Cheng, C. K., Yi, M., Lui, K. O. & Huang, Y. Targeting endothelial dysfunction and inflammation. *J. Mol. Cell. Cardiol.* **168**, 58–67. https://doi.org/10.1016/j.yjmcc.2022.04.011 (2022).
47. Gao, H., Wu, X., Zhang, Y., Liu, G. & Zhang, X. Novel predictive factor for erectile dysfunction: systemic immune inflammation index. *Int. J. Impot. Res.* https://doi.org/10.1038/s41443-024-00969-5 (2024).
48. Chen, D. et al. Association between the systemic immune-inflammation index and erectile dysfunction: a cross-sectional study. *Immun. Inflamm. Dis.* **12**, e1363 (2024).
49. Fujita, N. et al. Association between Advanced Glycation End-Products, carotenoids, and severe Erectile Dysfunction. *World J. Mens Health*. **41**, 701–711. https://doi.org/10.5534/wjmh.220154 (2023).
50. Crafa, A. et al. Mechanisms suggesting a relationship between Vitamin D and Erectile Dysfunction: an overview. *Biomolecules*. **13** https://doi.org/10.3390/biom13060930 (2023).
51. Holick, M. F. et al. Evaluation, treatment, and prevention of vitamin D deficiency: an endocrine Society clinical practice guideline. *J. Clin. Endocrinol. Metab.* **96**, 1911–1930. https://doi.org/10.1210/jc.2011-0385 (2011).
52. Traish, A. M., Goldstein, I. & Kim, N. N. Testosterone and erectile function: from basic research to a new clinical paradigm for managing men with androgen insufficiency and erectile dysfunction. *Eur. Urol.* **52**, 54–70. https://doi.org/10.1016/j.eururo.2007.02.034 (2007).
53. Armagan, A., Kim, N. N., Goldstein, I. & Traish, A. M. Dose-response relationship between testosterone and erectile function: evidence for the existence of a critical threshold. *J. Androl.* **27**, 517–526. https://doi.org/10.2164/jandrol.05157 (2006).
54. Lugg, J. A., Rajfer, J. & González-Cadavid, N. F. Dihydrotestosterone is the active androgen in the maintenance of nitric oxide-mediated penile erection in the rat. *Endocrinology*. **136**, 1495–1501. https://doi.org/10.1210/endo.136.4.7534702 (1995).
55. Nieschlag, E. et al. Investigation, treatment and monitoring of late-onset hypogonadism in males. ISA, ISSAM, and EAU recommendations. *Eur. Urol.* **48**, 1–4. https://doi.org/10.1016/j.eururo.2005.04.027 (2005).
56. Mulhall, J. P. et al. Evaluation and management of Testosterone Deficiency: AUA Guideline. *J. Urol.* **200**, 423–432. https://doi.org/10.1016/j.juro.2018.03.115 (2018).
57. in EAU Annual Congress Barcelona.
58. Duarte, C. K. & de Abreu Silva, L. Almeida Santos, A. Measuring muscle mass: mid-arm muscle circumference or mid-upper arm circumference? *Clin. Nutr.* **41**, 2424–2425 (2022). de.
59. Zhuang, C. L. et al. Reference values of low body mass index, mid-upper arm circumference, and calf circumference in cancer patients: a nationwide multicenter observational study. Nutrition (Burbank, Los Angeles County. *Calif)*. 99–100. https://doi.org/10.1016/j.nut.2022.111688 (2022).
60. Surmeli, D. M. et al. Association between Sarcopenia and erectile dysfunction in older males. *Arch. Gerontol. Geriatr.* **99**, 104619. https://doi.org/10.1016/j.archger.2021.104619 (2022). https://doi.org:.
61. Uçak, S. et al. Association between Sarcopenia and erectile dysfunction in males with type II diabetes mellitus. *Aging Male*. **22**, 20–27. https://doi.org/10.1080/13685538.2018.1441276 (2019).
62. Siegel, T. et al. The development of erectile dysfunction in men treated for prostate cancer. *J. Urol.* **165**, 430–435. https://doi.org/10.1097/00005392-200102000-00019 (2001).
63. Cellek, S., Cameron, N. E., Cotter, M. A., Fry, C. H. & Ilo, D. Microvascular dysfunction and efficacy of PDE5 inhibitors in BPH-LUTS. Nature reviews. *Urology*. **11**, 231–241. https://doi.org/10.1038/nrurol.2014.53 (2014).
64. Andersson, K. E. et al. Tadalafil for the treatment of lower urinary tract symptoms secondary to benign prostatic hyperplasia: pathophysiology and mechanism(s) of action. *Neurourol. Urodyn.* **30**, 292–301. https://doi.org/10.1002/nau.20999 (2011).
65. Stief, C. G., Porst, H., Neuser, D., Beneke, M. & Ulbrich, E. A randomised, placebo-controlled study to assess the efficacy of twice-daily vardenafil in the treatment of lower urinary tract symptoms secondary to benign prostatic hyperplasia. *Eur. Urol.* **53**, 1236–1244. https://doi.org/10.1016/j.eururo.2008.01.075 (2008).

66. Gacci, M. et al. A systematic review and meta-analysis on the use of phosphodiesterase 5 inhibitors alone or in combination with α-blockers for lower urinary tract symptoms due to benign prostatic hyperplasia. *Eur. Urol.* **61**, 994–1003. https://doi.org/10.1016/j.eururo.2012.02.033 (2012).
67. Braun, M. H. et al. Lower urinary tract symptoms and erectile dysfunction: co-morbidity or typical Aging Male symptoms? Results of the Cologne Male Survey. *Eur. Urol.* **44**, 588–594. https://doi.org/10.1016/s0302-2838(03)00358-0 (2003).
68. Echeverri Tirado, L. C., Ferrer, J. E. & Herrera, A. M. Aging and Erectile Dysfunction. *Sex. Med. Reviews.* **4**, 63–73. https://doi.org/10.1016/j.sxmr.2015.10.011 (2016).
69. Feldman, H. A., Goldstein, I., Hatzichristou, D. G., Krane, R. J. & McKinlay, J. B. Impotence and its medical and psychosocial correlates: results of the Massachusetts Male Aging Study. *J. Urol.* **151**, 54–61. https://doi.org/10.1016/S0022-5347(17)34871-1 (1994).
70. Jarow, J. P. Obesity and erectile dysfunction. *J. Urol.* **176**, 1282. https://doi.org/10.1016/j.juro.2006.06.135 (2006).
71. Almehmadi, Y., Yassin, D. J. & Yassin, A. A. Erectile dysfunction is a prognostic indicator of comorbidities in men with late onset hypogonadism. *Aging Male.* **18**, 186–194. https://doi.org/10.3109/13685538.2015.1046044 (2015).
72. Pizzol, D. et al. Associations between body mass index, waist circumference and erectile dysfunction: a systematic review and META-analysis. *Reviews Endocr. Metabolic Disorders.* **21**, 657–666. https://doi.org/10.1007/s11154-020-09541-0 (2020).
73. Fillo, J., Levcikova, M., Ondrusova, M., Breza, J. & Labas, P. Importance of different grades of abdominal obesity on Testosterone Level, Erectile Dysfunction, and clinical coincidence. *Am. J. Men's Health.* **11**, 240–245. https://doi.org/10.1177/1557988316642213 (2017).
74. Han, T. S. et al. Impaired quality of life and sexual function in overweight and obese men: the European Male Ageing Study. *Eur. J. Endocrinol.* **164**, 1003–1011. https://doi.org/10.1530/eje-10-1129 (2011).
75. Macdonald, E. J., Gaines, J. M., Kim, J. I. & Paduch, D. A. Exploring the relationship between socioeconomic status and erectile dysfunction: an analysis of the National Health and Nutrition Examination Survey. *Int. J. Impot. Res.* **35**, 478–483. https://doi.org/10.1038/s41443-022-00584-2 (2023).
76. Rosen, R. C. Psychogenic erectile dysfunction. Classification and management. *Urologic. Clin. North. Am.* **28**, 269–278. https://doi.org/10.1016/s0094-0143(05)70137-3 (2001).
77. Rasmussen, S., Balasubramaniam, K., Jarbøl, D. E., Søndergaard, J. & Haastrup, P. F. Socioeconomic status and barriers for contacting the general practitioner when bothered by erectile dysfunction: a population-based cross-sectional study. *BMC Fam. Pract.* **21**, 166. https://doi.org/10.1186/s12875-020-01238-2 (2020).
78. Cohen, S., Doyle, W. J. & Baum, A. Socioeconomic status is associated with stress hormones. *Psychosom. Med.* **68**, 414–420. https://doi.org/10.1097/01.psy.0000221236.37158.b9 (2006).
79. Fang, M. Trends in the prevalence of diabetes among U.S. adults: 1999–2016. *Am. J. Prev. Med.* **55**, 497–505. https://doi.org/10.1016/j.amepre.2018.05.018 (2018).
80. Fang, M. Trends in Diabetes Management among US adults: 1999–2016. *J. Gen. Intern. Med.* **35**, 1427–1434. https://doi.org/10.1007/s11606-019-05587-2 (2020).
81. Cignarelli, A. et al. Role of glucose-lowering medications in Erectile Dysfunction. *J. Clin. Med.* **10** https://doi.org/10.3390/jcm10112501 (2021).
82. Nguyen, H. M. T., Gabrielson, A. T. & Hellstrom, W. J. G. Erectile Dysfunction in Young Men-A Review of the prevalence and risk factors. *Sex. Med. Reviews.* **5**, 508–520. https://doi.org/10.1016/j.sxmr.2017.05.004 (2017).

Author contributions

Xing-Yu Chen and Xin Qin contributed to the study conception and design. Material preparation, data collection and preprocessing were performed by Wen-Ting Lu and Di Zhang. Model training and analysis were performed by Xing-Yu Chen and Mo-Yao Tan. The first draft of the manuscript was written by Xing-Yu Chen and proofread by Di Zhang. All authors read and approved the final manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Ethics approval and consent to participate

This study was not subject to ethical review and approval procedures as it utilized data from the publicly available National Health and Nutrition Examination Survey (NHANES) database. The use of NHANES data was authorized by the National Center for Health Statistics (NCHS), USA. Ethical approval for the study protocols (NHANES 2001–2004) was obtained from the NCHS Research Ethics Review Committee, and informed consent was provided by all NHANES survey participants.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-78797-2.

Correspondence and requests for materials should be addressed to X.Q.