Heliyon 10 (2024) e38124

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Determinants and prediction of hypertension among Chinese middle-aged and elderly adults with diabetes: A machine learning approach

Lijun Mao [a], Luotao Lin [b], Zumin Shi [c], Hualing Song [a], Hailei Zhao [a,**], Xianglong Xu [a,d,e,*]

[a] School of Public Health, Shanghai University of Traditional Chinese Medicine, Shanghai, China
[b] Nutrition and Dietetics Program, Department of Individual, Family, and Community Education, University of New Mexico, United States
[c] Human Nutrition Department, College of Health Sciences, QU Health, Qatar University, Qatar
[d] School of Translational Medicine, Faculty of Medicine, Nursing and Health Sciences, Monash University, Melbourne, Victoria, Australia
[e] Artificial Intelligence and Modelling in Epidemiology Program, Melbourne Sexual Health Centre, Alfred Health, Carlton, Victoria, Australia

ARTICLE INFO

Keywords:
Multimorbidity
Hypertension
Diabetes
China
Middle-aged and elderly adults
Machine learning
Prediction model

ABSTRACT

Objective: Multimorbidity, particularly diabetes combined with hypertension (DCH), is a significant public health concern. Currently, there is a gap in research utilizing machine learning (ML) algorithms to predict hypertension risk in Chinese middle-aged and elderly diabetic patients, and gender differences in DCH comorbidity patterns remain unclear. We aimed to use ML algorithms to predict DCH and identify its determinants among middle-aged and elderly diabetic patients in China.

Study design: Cross-sectional study.

Methods: Data were collected on 2775 adults with diabetes aged ≥45 years from the 2015 China Health and Retirement Longitudinal Study. We employed nine ML algorithms to develop prediction models for DCH. The performance of these models was evaluated using the area under the curve (AUC). Additionally, we conducted variable importance analysis to identify key determinants.

Results: Our results showed that the best prediction models for the overall population, men, and women were extreme gradient boosting (AUC = 0.728), light gradient boosting machine (AUC = 0.734), and random forest (AUC = 0.737), respectively. Age, waist circumference, body mass index, creatinine level, triglycerides, taking Western medicine, high-density lipoprotein cholesterol, blood urea nitrogen, total cholesterol, low-density lipoprotein cholesterol, and sleep disorders were identified as common important predictors by all three populations.

Conclusions: ML algorithms showed accurate predictive capabilities for DCH. Overall, non-linear ML models outperformed traditional logistic regression for predicting DCH. DCH predictions exhibited variations in predictors and model accuracy by gender. These findings could help identify DCH early and inform the development of personalized intervention strategies.

* Corresponding author. School of Public Health, Shanghai University of Traditional Chinese Medicine 1200 Cai Lun Road, Shanghai, 201203, China.
** Corresponding author.
E-mail address: xianglongxu@shutcm.edu.cn (X. Xu).

https://doi.org/10.1016/j.heliyon.2024.e38124
Received 3 July 2024; Received in revised form 7 September 2024; Accepted 18 September 2024
Available online 19 September 2024

1. Introduction

Diabetes and hypertension are prevalent chronic diseases worldwide, with an increasing incidence. The International Diabetes Federation anticipates a global prevalence of diabetes to reach 12.2 % (783.7 million) by 2045 [1]. From 1990 to 2019, the number of hypertension cases surged from 650 million to 1.3 billion globally [2]. Approximately 1.28 billion adults aged 30 to 79 suffer from hypertension worldwide [3]. China has the highest number of diabetic patients globally [4]. The prevalence and risk of diabetes escalate significantly after the age of 45, and this trend is expected to continue as the population ages [5,6]. It is estimated that over 50 % of diabetics also have hypertension, which accelerates the progression of nephropathy and worsens diabetes symptoms [7]. Diabetes combined with hypertension (DCH) significantly elevates the risk of cardiovascular diseases and poses a substantial socioeconomic burden [8]. Hence, early identification and intervention of DCH is essential to improve the prognosis of patients with diabetes and reduce the incidence of serious complications.

One of the global targets for non-communicable diseases is to reduce the prevalence of hypertension by 33 % between 2010 and 2030 [3]. Efficiently controlling hypertension among middle-aged and elderly adults with diabetes and accurately predicting DCH are crucial for advancing toward this goal. In recent years, the advent of artificial intelligence has spurred the widespread application of machine learning (ML) algorithms in clinical research [9]. Compared to traditional statistical methods, ML algorithms excel at handling intricate nonlinear relationships, interactions, and multiple covariates, significantly bolstering the predictive power of the models [10]. As such, ML algorithms can be utilized to improve prediction accuracy and support early intervention strategies, thereby effectively managing hypertension in diabetic patients.

Despite this, there is still an absence of DCH prediction models for Chinese middle-aged and elderly diabetic patients. Previous studies have predominantly employed logistic regression (LR) models and Cox regression methods to construct prediction models for DCH, demonstrating good predictive performance [11–15]. Furthermore, some studies have focused on exploring predictors [16–18], while others have reported using ML algorithms to predict DCH. For instance, a study in the United Arab Emirates successfully predicted eight diabetic complications, including hypertension, using the random forest model (AUC = 0.736) [19]. Moreover, another study involving 164 diabetic patients in Massachusetts, United States of America, utilized liquid biomarkers to predict DCH, with the linear discriminant analysis classifier achieving the highest accuracy rate of 61.2 % [20]. Additionally, a study based on 2080 diabetic patients from Qingdao, China, established a DCH prediction model incorporating blood pressure as a predictor, with an AUC exceeding 0.9 [21]. Among these, two studies [19,20] were conducted in other countries, and one focused on Chinese individuals aged 12–83 years [21]. However, there has not been an ML algorithm prediction study of DCH covering the characteristics of Chinese middle-aged and elderly adults. Additionally, research investigating gender differences in both predictors and predictive accuracy of DCH models is scarce. Existing evidence suggests that men generally have higher blood pressure levels than age-matched premenopausal women, while women often experience an increase in blood pressure after menopause [22,23]. This gender-based variation warrants further exploration to enhance the precision and applicability of DCH prediction models.

Our study aimed to develop predictive models for DCH using nine ML algorithms: LR, adaptive boosting (AdaBoost), gradient boosting machine (GBM), gaussian naive bayes (GNB), light gradient boosting machine (LGBM), RF, support vector machine (SVM), k-nearest neighbor classification (KNN), and extreme gradient boosting (XGBoost). Additionally, we aimed to identify key factors influencing DCH and explore variations in model performance and significant predictors among different gender groups. These findings have the potential to deepen our insights into DCH pathogenesis, while also informing gender-specific personalized medical recommendations and hypertension management strategies for diabetics, ultimately enhancing the quality of life for diabetics.

2. Methods

2.1. Study design

Participants were selected from the China Health and Retirement Longitudinal Study (CHARLS) Phase III (2015) to conduct a cross-sectional analysis. CHARLS is a nationally representative longitudinal survey that focuses on middle-aged and older adults in China, covering 150 county-level units, 450 village-level units, and approximately 10,000 households across the country [24]. In our study, the dataset consisted of three components: self-reported questionnaire, physical examination, and blood biomarkers. The inclusion criteria for participants were as follows: (1) aged 45 years or older; (2) physician-diagnosed or self-reported cases of diabetes. Conversely, individuals with missing information regarding the prevalence of hypertension were excluded. Ultimately, 2775 diabetic patients were included in our analysis and modeling(Fig. 1).

2.2. Predictors

Informed by expert insights and literature reviews [11,12,15], we included 23 predictors collected through questionnaires or measurements. The self-reported questionnaire encompassed: (1) demographic factors, such as gender, age, education level, region (east, middle, or west), and residence (urban or rural); (2) lifestyle factors, including smoking and drinking; (3) health status, including depression, which was evaluated through the administration of the 10-item Centre for Epidemiological Studies Depression Scale (CES-D-10) questionnaire [25], sleep disorders, and body pain; and (4) treatment information, detailing whether participants used traditional Chinese medicine, Western medicine, or insulin for diabetes management. Additionally, the measurements included: (1) physical examination, including body mass index (BMI) and waist circumference; and (2) blood biomarkers, including fasting blood

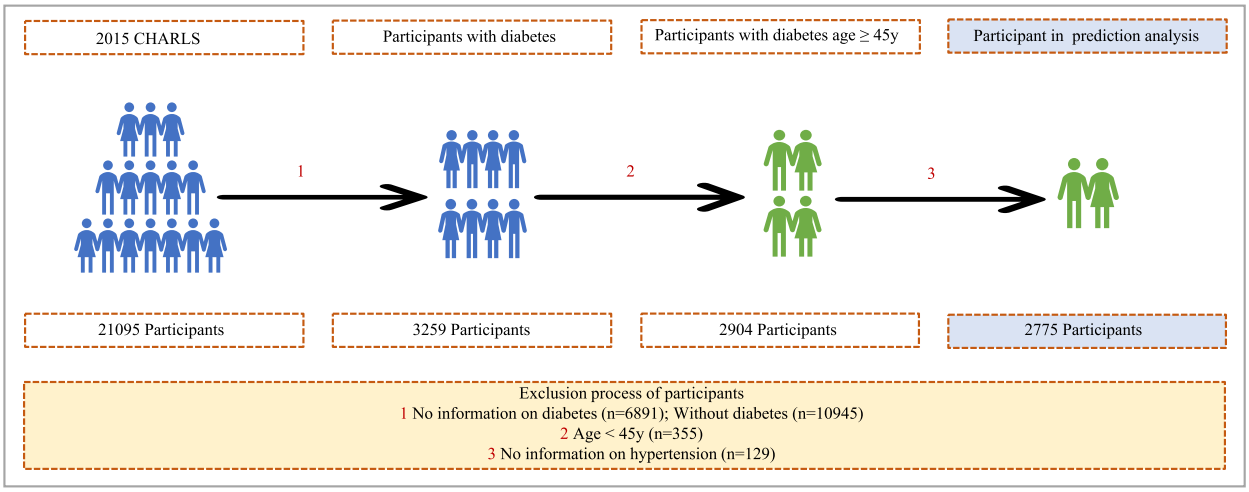

Fig. 1. Flowchart of this study.

glucose, glycated hemoglobin, high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C), total cholesterol, creatinine level, triglycerides, and blood urea nitrogen.

2.3. Definition of diabetes

In our study, the diagnosis of diabetes was established if any one of the following three criteria was met: (1) a previously physician-confirmed diagnosis of diabetes or self-reported diabetes; (2) a fasting blood glucose level of ≥7.0 mmol/L and/or a glycated hemoglobin level of ≥6.5 % [26]; (3) a positive response to the question, "Are you currently taking any of the following treatments to treat or control your diabetes? Utilizing traditional Chinese medicine, utilizing Western medicine, or administering insulin injections".

2.4. Measurement of outcome

The outcome was whether diabetic patients had comorbid hypertension. Hypertension was diagnosed based on any of the following three criteria: (1) a confirmed physician diagnosis or self-reported hypertension; (2) a diastolic blood pressure of ≥90 mmHg and/or a systolic blood pressure of ≥140 mmHg [27]; (3) affirmatively responding to the inquiry, "Are you currently undergoing any of the following treatments to manage your hypertension? Utilizing traditional Chinese medicine or Western medicine".

2.5. Analysis and modeling

Statistical analysis was performed using R 4.3.1, while model development was conducted in Python 3.8. Continuous variables were summarized using the median and interquartile range, and between-group comparisons were conducted using the Mann-Whitney *U* test. Categorical variables were summarized by presenting the frequency (n) and percentage (%). We used chi-square or Fisher's exact test to analyze group differences. The significance level was set at a p-value of less than 0.05. For all variables with a missing rate below 20 %, we employed two strategies to handle missing data: imputation with placeholders and interpolation using RF. Subsequently, we trained models using datasets processed by these two methods. Experimental results showed that datasets filled with placeholders outperformed those interpolated using RF in model training. Therefore, to preserve the distribution of the original data, we chose to fill in missing values with designated placeholders to represent unavailable information. During model training, missing data was not categorized. After preprocessing, we randomly split the dataset into 70 % training and 30 % testing sets using the *createDataPartition* function from R's *caret* package, stratified by the Outcome variable. To ensure reproducibility, a random seed was set, and the dataset rows were shuffled using the sample function before splitting. After splitting, a statistical analysis confirmed that there were no significant differences between the sets (p-value >0.05), indicating an even distribution. According to the no-free-lunch theorem [28], algorithms do not consistently demonstrate superiority or inferiority across all possible datasets. Therefore, we leveraged Python's *sklearn* and *XGBoost* packages to establish a series of commonly used ML models: LR, AdaBoost, GBM, GNB, LGBM, RF, SVM, KNN, and XGBoost. AdaBoost iteratively trains weak classifiers and combines them into a strong classifier with high accuracy, noise resistance, and the ability to handle imbalanced datasets [29]. GBM utilizes decision trees as weak learners and iteratively enhances their predictive performance [30]. GNB is recognized for its fast computation, low memory usage, adaptability, and robustness [31]. LGBM is a scalable gradient boosting framework suitable for high-dimensional features [32]. RF is proficient in handling high-dimensional data with strong classification and regression capabilities [33]. SVM excels in identifying nonlinear and high-dimensional patterns [34]. KNN makes classification decisions by measuring distances between data points [35]. In our research, we selected LR as the baseline model. We employed Bayesian optimization combined with 5-fold internal cross-validation to search for

the optimal combination of hyperparameters. The discriminatory capacity of the models was assessed by constructing receiver operating characteristic (ROC) curves and calculating the AUC. AUCs falling between 0.7 and 0.8 were considered acceptable, those between 0.8 and 0.9 were regarded as good, and values exceeding 0.9 were classified as excellent [36]. Additionally, we conducted variable importance analysis using XGBoost, which automatically deals with interaction features [37]. This allowed us to quantitatively assess the contribution of each feature to the model, thereby enhancing its interpretive capacity.

3. Results

3.1. Characteristics of the study participants

The study included a total of 2775 participants, with 63.6 % being DCH (Table 1). Among the DCH individuals, women accounted for 56.6 % and men for 43.4 %. The median age was 65 years, with an interquartile range of 58–71 years. Men generally had higher levels of education than women, with 28.5 % of men having less than primary school education compared to 59.0 % of women. A detailed analysis of differences between subpopulations of men and women in (Table 2).

3.2. Comparison of model performance

Six of the nine ML algorithms demonstrated acceptable prediction performance. These algorithms included GBM (AUC = 0.705), LGBM (AUC = 0.719), AdaBoost (AUC = 0.720), RF (AUC = 0.723), KNN (AUC = 0.726), and XGBoost (AUC = 0.729). In subgroup analyses, among men, six algorithms showed acceptable prediction capabilities: KNN (AUC = 0.706), GBM (AUC = 0.702), RF (AUC = 0.714), XGBoost (AUC = 0.711), AdaBoost (AUC = 0.724), and LGBM (AUC = 0.734). Among women, eight algorithms demonstrated acceptable prediction performance: SVM (AUC = 0.711), KNN (AUC = 0.712), LR (AUC = 0.713), LGBM (AUC = 0.725), GBM (AUC = 0.730), XGBoost (AUC = 0.733), AdaBoost (AUC = 0.735), and RF (AUC = 0.737). In all three groups, AdaBoost, LGBM, GBM, KNN, RF, and XGBoost exhibited acceptable AUCs (Fig. 2). The hyperparameters used during model training were outlined (Table 3), and an evaluation of the prediction accuracy of ML models on the testing dataset was provided (Table 4)

3.3. Determinants of DCH

Drawing upon models exhibiting acceptable AUCs across all three groups, we consolidated the top 10 important predictors identified by the underlying classifiers of these algorithms into a union set (Table 5). The results revealed that age, waist circumference, BMI, creatinine level, triglycerides, HDL-C, blood urea nitrogen, taking Western medicine, sleep disorders, total cholesterol, and LDL-C were common predictive factors across all groups. Notably, education level, drinking, and smoking were unique to men, while depression and body pain were specific to women. To understand the contributions of these predictors within a high-performance tree-based model, we utilized XGBoost to rank the variable importance (Fig. 3.). The ranked predictors were highly consistent with the union set. In the overall population, age emerged as the highest-ranked variable, followed by BMI and waist circumference. BMI was the highest-ranked variable among men but not among women. Additionally, education level and drinking were also identified as unique features of men, while depression stood out as a unique characteristic among women.

4. Discussion

We used ML algorithms to predict DCH in middle-aged and elderly Chinese diabetic patients, based on self-reported questions, physical examination, and blood biomarker data. Our findings showed that ML algorithms could accurately predict DCH among middle-aged and elderly adults. Our study also investigated gender-specific variations in predictors and model performance for predicting DCH. We found that there were gender-specific variations in predictors and model accuracy.

Our results demonstrated that the AdaBoost, RF, KNN, GBM, LGBM, and XGBoost models surpassed the LR model in accurately predicting DCH. Overall, non-linear models outperformed linear models, suggesting that non-linear models will be more effective in identifying DCH in clinical settings. This contrasts sharply with Jadhav's study [20], which relied solely on biomarker data and did not fully leverage the capabilities of nonlinear models in handling more complex datasets. Our study emphasizes the advantages of non-linear models. However, our exploration did not include ensemble modeling, a technique that combines predictions from various models to potentially enhance predictive performance [38]. Whether ensemble models are superior and whether the models presented in our study are effective in clinical practice requires further validation. Although the current study was conducted in a Chinese population, similar model development approaches can provide valuable technical insights for predicting DCH across different ethnic groups.

Our study had several advantages. Firstly, in terms of predictors, we acknowledged the significant impact of hypoglycemic medication on the development of hypertension in diabetic patients, which has been overlooked in previous DCH prediction models [19–21]. Our dataset included the use of hypoglycemic medications, categorized into traditional Chinese medicine, Western medicine, and insulin. Moreover, we conducted separate evaluations to assess the impacts of these treatments on hypertension. This consideration not only provides more comprehensive information but also has the potential to guide the development of personalized and

Table 1
Characteristics of the study participants with diabetes (n = 2775).

Characteristic	Overall		Men		Women	
	Hypertension	Non-hypertension	Hypertension	Non-hypertension	Hypertension	Non-hypertension
	N = 1766[a]	N = 1009[a]	N = 766[a]	N = 459[a]	N = 999[a]	N = 550[a]
Gender						
Men	766(43.4)	459(45.5)				
Women	999(56.6)	550(54.5)				
Missing	1(0.0)					
Age (years)	65 (58,71)	61 (54,67)	65 (58,71)	61 (55,68)	65 (58,71)	60 (53,66)
Education level						
Less than elementary school	806(45.6)	437(43.4)	217(28.5)	132(28.7)	589(59.0)	305(55.5)
Elementary school	410(23.2)	219(21.7)	217(28.3)	117(25.5)	193(19.3)	102(18.5)
Middle school	325(18.5)	229(22.7)	187(24.4)	130(28.3)	138(13.8)	99(18.0)
High school or above	210(11.9)	107(10.6)	140(18.2)	73(16.0)	69(6.9)	34(6.2)
Missing	15(0.8)	17(1.6)	5(0.6)	7(1.5)	10(1.0)	10(1.8)
Region						
East	582(33.0)	323(32.0)	261(34.1)	138(30.1)	321(32.1)	185(33.6)
Midland	714(40.4)	393(38.9)	289(37.7)	176(38.3)	424(42.4)	217(39.5)
West	470(26.6)	293(29.0)	216(28.2)	145(31.6)	254(25.4)	148(26.9)
Residence						
Urban	523(29.7)	248(24.6)	234(30.6)	119(25.9)	289(29.0)	129(23.5)
Rural	1236(70.0)	759(75.2)	530(69.2)	339(73.9)	706(70.6)	420(76.4)
Missing	7(0.3)	2(0.2)	2(0.2)	1(0.2)	4(0.4)	1(0.1)
Smoking						
No	1082(61.2)	597(59.2)	172(22.4)	92(20.1)	910(91.1)	505(91.8)
Yes	682(38.7)	411(40.8)	593(77.5)	366(79.7)	88(8.8)	45(8.2)
Missing	2(0.1)	1(0.0)	1(0.1)	1(0.2)	1(0.1)	
Drinking						
No	1252(70.9)	695(68.9)	381(49.7)	223(48.6)	870(87.1)	472(85.8)
Yes	513(29.1)	314(31.1)	385(50.3)	236(51.4)	128(12.8)	78(14.2)
Missing	1(0.0)				1(0.1)	
Depression						
No	848(48.0)	554(54.9)	420(54.8)	282(61.4)	427(42.8)	272(49.5)
Yes	639(36.2)	365(36,2)	246(32.1)	141(30.7)	393(39.3)	224(40.7)
Missing	279(15.8)	90(8.9)	100(13.1)	36(7.9)	179(17.9)	54(9.8)
Body pain						
No	1067(60.4)	687(68.0)	532(69.5)	331(72.1)	534(53.5)	356(64.7)
Yes	600(34.0)	301(30.0)	188(24.5)	118(25.7)	412(41.2)	183(25.1)
Missing	99(5.6)	21(2.0)	46(6.0)	10(2.2)	53(5.3)	11(0.2)
Sleep disorders						
No	751(42.5)	401(39.7)	262(34.2)	157(34.2)	489(49.0)	244(44.4)
Yes	883(50.0)	559(55.4)	449(58.6)	285(62.1)	433(43.3)	274(49.8)
Missing	132(7.5)	49(4.9)	55(7.2)	17(3.7)	77(7.7)	32(5.8)
Taking traditional Chinese medicine						
No	1643(93.0)	963(95.4)	709(92.6)	432(94.1)	933(93.4)	531(96.5)
Yes	123(7.0)	46(4.6)	57(7.4)	27(5.9)	66(6.6)	19(3.5)
Taking Western medicine						
No	1163(65.9)	785(77.8)	512(66.8)	370(80.6)	650(65.1)	415(75.5)
Yes	603(34.1)	224(22.2)	254(33.2)	89(19.4)	349(34.9)	135(24.5)
Taking insulin injections						
No	1598(90.5)	955(94.6)	686(89.6)	432(94.1)	911(91.2)	523(95.1)
Yes	168(9.5)	54(5.4)	80(10.4)	27(5.9)	88(8.8)	27(4.9)
Body mass index						
Underweight	45(2.5)	61(6.1)	37(8.0)	22(2.2)	24(4.3)	37(8.0)
Normal weight	452(25.6)	453(44.9)	226(49.2)	240(24.0)	227(41.3)	226(49.2)
Overweight	658(37.2)	349(34.6)	146(31.8)	374(37.5)	203(36.9)	146(31.8)
Obese	402(22.8)	136(13.5)	46(10.0)	260(26.0)	90(16.5)	46(10.0)
Missing	209(11.9)	10(0.9)	4(1.0)	103(10.3)	6(1.0)	4(1.0)
Waist circumference (cm)	93 (86,99)	87 (80,94)	86 (79,94)	93 (86,99)	88 (81,95)	86 (79,94)
Missing	204(11.5)	10(0.9)	3(0.6)	100(10.0)	7(0.1)	3(0.6)
Glycated hemoglobin (%)	6.70 (6.10,7.60)	6.60 (6.00,7.30)	6.60 (6.00,7.50)	6.50 (5.90,7.30)	6.70 (6.20,7.70)	6.60 (6.10,7.30)
Missing	276(15.6)	66(6.5)	130(16.9)	26(5.0)	146(14.6)	40(7.3)
Fasting blood glucose(mmol/L)	7.1 (5.9,8.5)	7.1 (5.7,8.4)	7.2 (5.9,8.6)	7.2 (5.8,8.6)	7.0 (5.9,8.5)	7.0 (5.6,8.4)
Missing	278(15.6)	65(6.4)	131(17.1)	26(5.0)	147(14.7)	39(7.1)
Total Cholesterol(mmol/L)	4.87 (4.26,5.54)	4.77 (4.11,5.49)	4.58 (4.07,5.24)	4.53 (3.95,5.26)	5.06 (4.46,5.77)	5.01 (4.32,5.66)
Missing	281(15.9)	67(6.5)	133(17.3)	27(5.8)	148(14.8)	40(7.3)

(*continued on next page*)

Table 1 (*continued*)

Characteristic	Overall		Men		Women	
	Hypertension	Non-hypertension	Hypertension	Non-hypertension	Hypertension	Non-hypertension
	N = 1766[a]	N = 1009[a]	N = 766[a]	N = 459[a]	N = 999[a]	N = 550[a]
Triglycerides(mmol/L)	1.69 (1.20,2.48)	1.55 (1.05,2.33)	1.55 (1.10,2.34)	1.42 (0.95,2.18)	1.79 (1.31,2.60)	1.64 (1.13,2.40)
Missing	281(15.9)	67(6.5)	133(17.3)	27(5.8)	148(14.8)	40(7.3)
High-density lipoprotein cholesterol (mmol/L)	1.20 (1.05,1.39)	1.26 (1.07,1.44)	1.14 (1.01,1.33)	1.20 (1.03,1.40)	1.26 (1.11,1.42)	1.29 (1.12,1.46)
Missing	281(15.9)	67(6.5)	133(17.3)	27(5.8)	148(14.8)	40(7.3)
Low-density lipoprotein cholesterol (mmol/L)	2.69 (2.20,3.20)	2.62 (2.12,3.15)	2.52 (2.05,2.99)	2.49 (2.05,3.02)	2.82 (2.28,3.36)	2.77 (2.19,3.26)
Missing	281(15.9)	67(6.5)	133(17.3)	27(5.8)	148(14.8)	40(7.3)
Blood urea nitrogen(mmol/L)	5.40 (4.50,6.60)	5.40 (4.50,6.50)	5.50 (4.60,6.80)	5.60 (4.70,6.70)	5.30 (4.50,6.50)	5.10 (4.20,6.30)
Missing	281(15.9)	67(6.5)	133(17.3)	27(5.8)	148(14.8)	40(7.3)
Creatinine(mmol/L)	67 (57,82)	65 (56,78)	78 (68,90)	76 (66,86)	60 (53,69)	57 (51,66)
Missing	281(15.9)	68(6.5)	133(17.3)	27(5.8)	148(14.8)	41(7.4)

Note.
[a] n(%); Median (P_{25},P_{75}).

precise prevention strategies. Secondly, unlike the previous study [21], our research did not incorporate blood pressure data, enabling an examination of hypertension-related factors independent of blood pressure. This approach fosters a deeper insight into the mechanisms underlying hypertension. Thirdly, subgroup analyses focus on prediction models for different gender groups, while also elucidating gender-specific differences in predictors. As expected, the LGBM demonstrated superior predictive performance for men, while the RF model performed better for women. However, previous studies have not addressed potential gender-based differences in prediction.

We found that age, waist circumference, BMI, creatinine level, triglycerides, HDL-C, blood urea nitrogen, taking Western medicine, sleep disorders, total cholesterol, and LDL-C were important predictors of DCH. These findings align with previous studies [11,13,15], where age, BMI, waist circumference, creatinine level, and hypoglycemic medication were associated with DCH. Future research could further explore the association and mechanism between hypoglycemic drugs and DCH. Guo et al. [39] found that lipid levels were strongly associated with the occurrence of hypertension. Specifically, abnormal blood lipid levels result in reduced nitric oxide synthesis, leading to impaired vasoconstriction and relaxation and abnormally high blood pressure [40]. Therefore, more emphasis should be placed on dietary management for diabetic patients to control lipid levels. Abnormal creatinine levels indicate the severity of kidney damage, which may be exacerbated by diabetes. Furthermore, renal dysfunction is intricately tied to both the onset and progression of hypertension [41]. Sleep disorders stimulate the sympathetic nervous system, increasing its activity. Overactivation of the sympathetic nervous system can lead to accelerated heart rate, vasoconstriction, and subsequent blood pressure elevation [42]. Multiple studies have shown that insufficient sleep duration increases the risk of hypertension [43,44]. Our analysis also revealed some novel findings: age emerged as the most important predictor in both the overall and women's populations, while it is the second most important in the men's subgroup. Despite there being no statistically significant age disparity between genders in our initial analysis, this discrepancy hints at an intricate interplay between age and diverse factors that may vary by gender. To unravel these complexities, future research could explore the interactions between age and other predictors, accounting for biological nuances such as hormonal differences, which may modulate the impact of age on DCH risk in a gender-specific manner. We hypothesize that one reason for this gender difference could be the younger baseline age of men in our study compared to women. Additionally, BMI was the most influential predictor of DCH among men, whereas it only ranked third in importance for women. This may be related to male-specific physiological characteristics and metabolic pathways, which increase men's sensitivity to weight-related metabolic risks [45,46]. In addition, men and women often face different societal expectations and pressures related to body image and weight management, which can influence their health behaviors and ultimately their risk for DCH. This suggests a greater need for focusing on men's weight management to prevent DCH. Furthermore, alcohol drinking and smoking were crucial modifiable predictors specific to men, whereas depression and body pain were significant modifiable predictors specific to women. Previous studies have shown that men consume alcohol and smoke at much higher rates than women [47], and women are more likely to develop psychiatric disorders and experience insomnia compared to men [48,49]. Our findings may contribute to better identification of high-risk populations for DCH and suggest the need for gender-specific intervention strategies. Men should prioritize controlling their alcohol consumption and smoking, as well as improving their weight management. Meanwhile, women should focus on mental health and pain management to effectively reduce the risk for DCH.

Table 2
Analysis of differences between subpopulations of men and women.

Characteristic	Men, N = 1225[a]	Women, N = 1549[a]	*p*-value[b]
Age (years)	63 (57,70)	63 (56,70)	0.13
Education level			<0.001
Less than elementary school	349(28.5)	894(57.7)	
Elementary school	334(27.3)	295(19.1)	
Middle school	317(25.9)	237(15.3)	
High school or above	213(17.4)	103(6.6)	
Missing	12(0.9)	20(1.3)	
Region			0.079
East	399(32.6)	506(32.6)	
Midland	465(37.9)	641(41.3)	
West	361(29.3)	402(26.1)	
Residence			0.5
Urban	869(71.0)	1126(72.7)	
Rural	353(28.8)	418(27.0)	
Missing	3(0.2)	5(0.3)	
Smoking			<0.001
No	264(21.5)	1415(91.3)	
Yes	959(78.3)	133(8.6)	
Missing	2(0.2)	1(<0.1)	
Drinking			<0.001
No	604(49.3)	1342(86.6)	
Yes	621(50.7)	206(13.2)	
Missing		1(<0.1)	
Depression			<0.001
No	702(57.3)	699(45.1)	
Yes	387(31.6)	617(38.9)	
Missing	136(11.1)	233(15.0)	
Body pain			<0.001
No	863(70.4)	890(57.5)	
Yes	306(25.0)	595(38.4)	
Missing	56(4.6)	64(4.1)	
Sleep disorders			<0.001
No	419(34.2)	733(47.3)	
Yes	806(65.8)	816(52.7)	
Taking traditional Chinese medicine			0.13
No	1141(93.1)	1464(94.5)	
Yes	84(6.9)	85(5.5)	
Taking Western medicine			0.063
No	882(72)	1065(68.8)	
Yes	343(28)	484(31.2)	
Taking insulin injections			0.2
No	1118(91.3)	1434(92.6)	
Yes	107(8.7)	115(7.4)	
Body mass index			<0.001
Underweight	60(4.9)	46(3.0)	
Normal weight	438(35.8)	467(30.2)	
Overweight	429(35.0)	577(37.2)	
Obese	188(15.3)	350(22.6)	
Missing	110(9.0)	109(7.0)	
Waist circumference (cm)	91(82, 98)	91(84, 98)	0.4
Missing	107(8.7)	107(6.9)	
Glycated hemoglobin (%)	6.6(6.0, 7.4)	6.7(6.2, 7.5)	<0.001
Missing	156(12.7)	186(12.0)	
Fasting blood glucose (mmol/L)	7.2(5.8, 8.6)	7.0(5.8, 8.4)	0.085
Missing	157(12.8)	186(12.0)	
Total Cholesterol (mmol/L)	4.5(4.0, 5.2)	5.0(4.4, 5.7)	<0.001
Missing	160(13.0)	188(12.1)	
Triglycerides (mmol/L)	1.5(1.0, 2.2)	1.7(1.2, 2.4)	<0.001
Missing	160(13.0)	188(12.1)	
High-density lipoprotein cholesterol (mmol/L)	1.1(1.0, 1.3)	1.2(1.1, 1.4)	<0.001
Missing	160(13.0)	188(12.1)	
Low-density lipoprotein cholesterol(mmol/L)	2.5(2.0, 3.0)	2.8(2.2, 3.3)	<0.001
Missing	160(13.0)	188(12.1)	
Blood urea nitrogen(mmol/L)	5.6(4.6, 6.8)	5.2(4.4, 6.4)	<0.001
Missing	160(13.0)	188(12.1)	
Creatinine(mmol/L)	77(67, 88)	59(52, 68)	<0.001
Missing	160(13.0)	189(12.2)	

Note.
[a] n(%); Median (P_{25},P_{75}).
[b] Pearson's Chi-squared test; Wilcoxon rank sum test; Fisher's exact test.

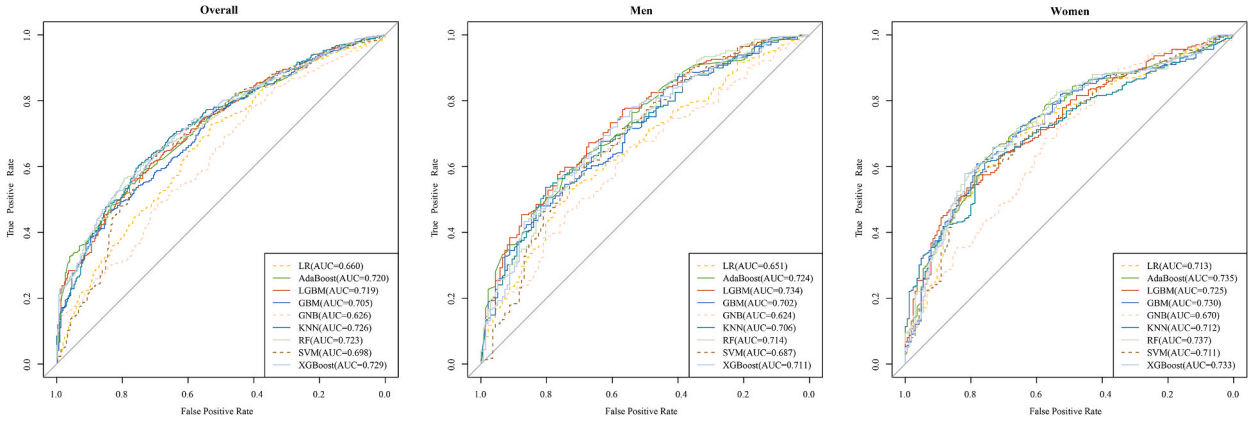

Fig. 2. Receiver operating characteristic curves for all models in the testing dataset. AUC, area under the curve; LR, logistic regression; AdaBoost, adaptive boosting; GBM, gradient boosting machine; GNB, gaussian naive Bayes; LGBM, light gradient boosting machine; RF, random forest; SVM, support vector machine; KNN, k-nearest neighbor classification; XGBoost, extreme gradient boosting.

Table 3
Summary of the hyperparameters of each model with acceptable AUCs.

Models	Parameters	Range	Parameter values		
			Overall	Men	Women
AdaBoost	learning_rate	[0.001,0.01,0.1,0.3,0.5,0.7,1.0]	0.01	0.1	0.1
	n_estimators	[10,50,500]	500	50	50
LGBM	learning_rate	[0.01,0.02,0.03,0.04,0.05,1]	0.03	0.01	0.01
	n_estimators	[100,200,500,1000]	100	100	200
	num_leaves	[10,50,100,200,300,400,500]	10	10	10
GBM	learning_rate	[0.001, 0.01,0.1,0.3,0.5,0.7,1.0]	0.1	0.001	0.01
	n_estimators	[50,100,150,300, 500]	100	50	300
	max_depth	[2,3,5,8,15,30]	3	3	2
KNN	n_neighbors	[20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30]	30	30	26
RF	n_estimators	[50, 75, 100, 125, 150, 200, 400]	400	75	400
	max_depth	[5, 10, 30, 60, 100]	10	5	5
	min_samples_leaf	[2,5,10]	10	8	10
	min_samples_split	[1,2,4,8,10]	2	5	2
XGBoost	learning_rate	[0.001, 0.01, 0.1, 0.3, 0.5, 0.7, 1.0]	0.01	0.01	0.001
	n_estimators	[100, 300, 500]	500	500	500
	max_depth	[5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20]	16	16	6

Note: AUC, area under the curve; LR, logistic regression; AdaBoost, adaptive boosting; GBM, gradient boosting machine; GNB, gaussian naive Bayes; LGBM, light gradient boosting machine; RF, random forest; SVM, support vector machine; KNN, k-nearest neighbor classification; XGBoost, extreme gradient boosting.

5. Limitations

Several limitations in our study must be acknowledged. Firstly, the cross-sectional design of our research limits our ability to draw causal inferences. Secondly, our study sample was obtained from a publicly available database that lacked dietary information and had more severe missing exercise-related variables. Previous research has demonstrated a strong correlation between diet, exercise, and the onset of hypertension [50,51]. Therefore, incorporating these factors could enhance the model's accuracy. Additionally, the absence of dietary information may have hindered a comprehensive evaluation of nutritional influences on study outcomes, potentially overlooking associations between diet and hypertension. Thirdly, there is room for enhancing the model's predictive accuracy (AUC = 0.73), necessitating further studies to refine it. Moreover, except for geographic data, all other pertinent information was self-reported or tested, making it susceptible to recall bias and measurement inaccuracies, which could affect the reliability of the model. Future studies should consider utilizing more objective measurements. Furthermore, our study is limited in its ability to differentiate between various types of diabetes due to the constraints of the data source. As a result, we are unable to accurately assess the specific risk profiles and implications of different types of diabetes. Lastly, the model's generalizability to other age groups or populations outside of China remains uncertain. Further external validation will be needed.

Table 4
Comparison of prediction performance of machine learning models on the testing dataset.

Groups	Models	AUC	Sensitivity	Specificity	Accuracy	F1 score
Overall	LR	0.660	0.768	0.440	0.560	0.559
	AdaBoost	0.720	0.745	0.571	0.634	0.597
	LGBM	0.719	0.772	0.550	0.631	0.603
	GBM	0.705	0.758	0.533	0.615	0.589
	GNB	0.626	0.752	0.367	0.507	0.525
	KNN	0.726	0.758	0.586	0.649	0.611
	RF	0.723	0.755	0.580	0.644	0.606
	SVM	0.698	0.781	0.495	0.599	0.586
	XGBoost	0.729	0.758	0.567	0.637	0.603
Men	LR	0.651	0.708	0.546	0.607	0.574
	AdaBoost	0.724	0.737	0.576	0.637	0.603
	LGBM	0.734	0.759	0.581	0.648	0.617
	GBM	0.702	0.715	0.537	0.604	0.575
	GNB	0.624	0.759	0.410	0.541	0.553
	KNN	0.706	0.766	0.563	0.639	0.614
	RF	0.714	0.759	0.568	0.639	0.612
	SVM	0.687	0.730	0.537	0.609	0.583
	XGBoost	0.711	0.781	0.524	0.620	0.606
Women	LR	0.713	0.733	0.605	0.651	0.599
	AdaBoost	0.735	0.752	0.615	0.664	0.614
	LGBM	0.725	0.727	0.589	0.638	0.588
	GBM	0.730	0.733	0.625	0.664	0.608
	GNB	0.670	0.739	0.441	0.547	0.537
	KNN	0.712	0.715	0.629	0.659	0.599
	RF	0.737	0.758	0.625	0.672	0.622
	SVM	0.711	0.733	0.602	0.649	0.598
	XGBoost	0.733	0.758	0.615	0.666	0.617

Note: AUC, area under the curve; LR, logistic regression; AdaBoost, adaptive boosting; GBM, gradient boosting machine; GNB, gaussian naive Bayes; LGBM, light gradient boosting machine; RF, random forest; SVM, support vector machine; KNN, k-nearest neighbor classification; XGBoost, extreme gradient boosting.

Table 5
The union set of the top 10 important predictors of machine learning models with acceptable AUCs in all three groups.

Groups	Predictors
Overall	age, body mass index, waist circumference, taking Western medicine, triglycerides, high-density lipoprotein cholesterol, creatinine level, fasting blood glucose, low-density lipoprotein cholesterol, total cholesterol, blood urea nitrogen, glycated hemoglobin, and sleep disorders.
Men	body mass index, age, high-density lipoprotein cholesterol, waist circumference, education level, drinking, blood urea nitrogen, taking Western medicine, total cholesterol, creatinine level, low-density lipoprotein cholesterol, fasting blood glucose, triglycerides, glycated hemoglobin, smoking, and sleep disorders.
Women	age, waist circumference, body mass index, creatinine level, depression, triglycerides, high-density lipoprotein cholesterol, blood urea nitrogen, taking Western medicine, sleep disorders, total cholesterol, low-density lipoprotein cholesterol, and body pain.

Note: AUC, area under the curve.

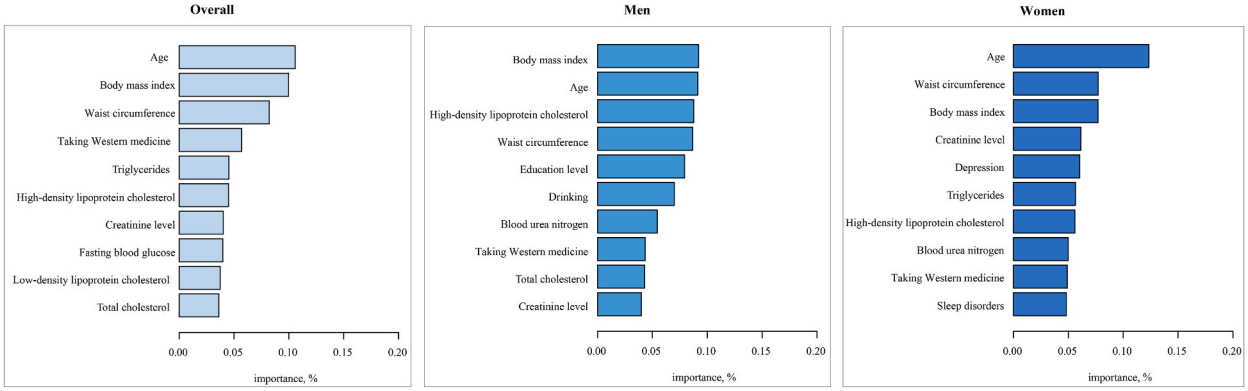

Fig. 3. The importance of the top 10 predictors in the prediction of DCH using extreme gradient boosting.

6. Conclusions

Our study demonstrates that ML algorithms could predict DCH. Overall, non-linear models showed better performance compared to traditional logistic regression for predicting DCH. Additionally, we explored gender-based variations in predictors and model accuracy, highlighting differences between men and women. Our prediction model and analysis of variable importance could offer valuable references for the early detection and health management of hypertension in middle-aged and elderly adults with diabetes in China.

Funding statement

This work was supported by the Traditional Chinese Medicine Research Project of the Shanghai Municipal Health Commission, grant number 2024QN108; Shanghai University of Traditional Chinese Medicine, grant number 2021LK008; and Shanghai University of Traditional Chinese Medicine, grant number KECJ2024019.

Ethical approval

The study was conducted in accordance with the Declaration of Helsinki. The Peking University Institutional Review Board granted ethical approval for all CHARLS waves. The Institutional Review Board (IRB) approval number for the self-reported questionnaire (incorporating physical examination) was IRB00001052-11015, while the approval number for biomarkers was IRB00001052-11014.

Data and code availability

The original data presented in the study are openly available in the China Health and Retirement Longitudinal Study at the URL: https://charls.pku.edu.cn/. The code can be obtained from the corresponding author upon a reasonable request.

CRediT authorship contribution statement

Lijun Mao: Writing – original draft, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Luotao Lin:** Writing – review & editing. **Zumin Shi:** Writing – review & editing. **Hualing Song:** Writing – review & editing, Funding acquisition. **Hailei Zhao:** Writing – review & editing, Supervision. **Xianglong Xu:** Writing – review & editing, Supervision, Software, Funding acquisition, Data curation, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:Xianglong Xu reports financial support was provided by Shanghai Municipal Health Commission. Hualing Song reports financial support was provided by Shanghai University of Traditional Chinese Medicine. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study used data from CHARLS. We would like to thank the CHARLS research team and each respondent for contributing to this study.

References

[1] H. Sun, P. Saeedi, S. Karuranga, M. Pinkepank, K. Ogurtsova, B.B. Duncan, C. Stein, A. Basit, J.C.N. Chan, J.C. Mbanya, M.E. Pavkov, A. Ramachandaran, S. H. Wild, S. James, W.H. Herman, P. Zhang, C. Bommer, S. Kuo, E.J. Boyko, D.J. Magliano, IDF Diabetes Atlas: global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045, Diabetes Res. Clin. Pract. 183 (2022) 109119, https://doi.org/10.1016/j.diabres.2021.109119.

[2] W.H. Organization, First WHO report details devastating impact of hypertension and ways to stop it (2023). https://www.who.int/news/item/19-09-2023-first-who-report-details-devastating-impact-of-hypertension-and-ways-to-stop-it. (Accessed 16 April 2024).

[3] W.H. Organization, Hypertension (2023). https://www.who.int/news-room/fact-sheets/detail/hypertension. (Accessed 16 April 2024).

[4] Y. Li, D. Teng, X. Shi, G. Qin, Y. Qin, H. Quan, B. Shi, H. Sun, J. Ba, B. Chen, J. Du, L. He, X. Lai, Y. Li, H. Chi, E. Liao, C. Liu, L. Liu, X. Tang, N. Tong, G. Wang, J. A. Zhang, Y. Wang, Y. Xue, L. Yan, J. Yang, L. Yang, Y. Yao, Z. Ye, Q. Zhang, L. Zhang, J. Zhu, M. Zhu, G. Ning, Y. Mu, J. Zhao, W. Teng, Z. Shan, Prevalence of diabetes recorded in mainland China using 2018 diagnostic criteria from the American Diabetes Association: national cross sectional study, Bmj 369 (2020) m997, https://doi.org/10.1136/bmj.m997.

[5] D.J. Magliano, E.J. Boyko, I.D.F.D.A.t.e.s. committee, IDF Diabetes Atlas, Idf diabetes atlas, International Diabetes Federation © International Diabetes Federation, Brussels, 2021, 2021.

[6] Collaborators G.B.D.D., Global, regional, and national burden of diabetes from 1990 to 2021, with projections of prevalence to 2050: a systematic analysis for the Global Burden of Disease Study 2021, Lancet 402 (10397) (2023) 203–234, https://doi.org/10.1016/s0140-6736(23)01301-6.

[7] S. Katayama, M. Hatano, M. Issiki, Clinical features and therapeutic perspectives on hypertension in diabetics, Hypertens. Res. 41 (4) (2018) 213–229, https://doi.org/10.1038/s41440-017-0001-5.

[8] D.I. Pavlou, S.A. Paschou, P. Anagnostis, M. Spartalis, E. Spartalis, A. Vryonidou, N. Tentolouris, G. Siasos, Hypertension in patients with type 2 diabetes mellitus: targets and management, Maturitas 112 (2018) 71–77, https://doi.org/10.1016/j.maturitas.2018.03.013.

[9] G.S. Handelman, H.K. Kok, R.V. Chandra, A.H. Razavi, M.J. Lee, H. Asadi, eDoctor: machine learning and the future of medicine, J. Intern. Med. 284 (6) (2018) 603–619, https://doi.org/10.1111/joim.12822.

[10] H.S.R. Rajula, G. Verlato, M. Manchia, N. Antonucci, V. Fanos, Comparison of conventional statistical methods with machine learning in medicine: diagnosis, drug development, and treatment, Medicina (Kaunas) 56 (9) (2020), https://doi.org/10.3390/medicina56090455.

[11] C.C. Lin, C.I. Li, C.S. Liu, C.H. Lin, M.C. Wang, S.Y. Yang, T.C. Li, A risk scoring system to predict the risk of new-onset hypertension among patients with type 2 diabetes, J. Clin. Hypertens. 23 (8) (2021) 1570–1580, https://doi.org/10.1111/jch.14322.

[12] J. Yang, X. Wang, S. Jiang, Development and validation of a nomogram model for individualized prediction of hypertension risk in patients with type 2 diabetes mellitus, Sci. Rep. 13 (1) (2023) 1298, https://doi.org/10.1038/s41598-023-28059-4.

[13] G. Sen, Y. Li, F. Sun, N. Deng, Risk factors of type 2 diabetes melitus complicating hypertension and its nomogram prediction model construction, Chongqing Med J 51 (13) (2022) 2189–2193+2198, https://doi.org/10.3969/j.issn.1671-8348.2022.13.006.

[14] H. Dan, C. Zhu, Q. Ni, Screening of key factors and risk prediction based on EMVS-logistic model: take diabetes comorbidity hypertension data as an example, J Lanzhou Arts & Sci Univ (NSE) 36 (4) (2022) 25–30, https://doi.org/10.13804/j.cnki.2095-6991.2022.04.023.

[15] H. Nong, H. Ning, X. Xu, L. Chen, Y. Lyu, H. Zhang, Establishment of a Nomogram prediction model for diabetes mellitus type 2 with hypertension, J. Guangxi Med. Univ. 40 (12) (2023) 2035–2042, https://doi.org/10.16190/j.cnki.45-1211/r.2023.12.016.

[16] V. Brar, N.B. Kandala, A. Terry, A. Thind, Predictors of multimorbidity (defined as diabetes + hypertension) amongst males aged 15-54 in India: an urban/rural split analysis, Int J Public Health 69 (2024) 1606660, https://doi.org/10.3389/ijph.2024.1606660.

[17] B. Dagnew, Y. Yeshaw, Predictors of isolated systolic hypertension among type 2 diabetes mellitus patients in Jimma University Specialized Hospital, Southwest Ethiopia, BMC Res. Notes 12 (1) (2019) 510, https://doi.org/10.1186/s13104-019-4550-3.

[18] A.S. Tegegne, Joint predictors of hypertension and type 2 diabetes among adults under treatment in amhara region (North-Western Ethiopia), Diabetes Metab Syndr Obes 14 (2021) 2453–2463, https://doi.org/10.2147/dmso.S309925.

[19] Y. Jian, M. Pasquier, A. Sagahyroon, F. Aloul, A machine learning approach to predicting diabetes complications, Healthcare (Basel) 9 (12) (2021), https://doi.org/10.3390/healthcare9121712.

[20] P. Jadhav, V. Selvaraju, S.P. Sathian, R. Swaminathan, Use of multiple fluid biomarkers for predicting the Co-occurrence of diabetes and hypertension using machine learning approaches, Annu Int Conf IEEE Eng Med Biol Soc 2023 (2023) 1–4, https://doi.org/10.1109/embc40787.2023.10340163.

[21] Y. Zhao, J. Han, X. Hu, B. Hu, H. Zhu, Y. Wang, X. Zhu, Hypertension risk prediction models for patients with diabetes based on machine learning approaches, Multimed Tools Appl 83 (20) (2024) 59085–59102, https://doi.org/10.1007/s11042-023-17926-x.

[22] J. Mehta, J.E. Manson, Menopausal hormone therapy and hypertension: minimizing risk, Menopause 28 (11) (2021) 1201–1202, https://doi.org/10.1097/gme.0000000000001868.

[23] L.A. Ramirez, J.C. Sullivan, Sex differences in hypertension: where we have been and where we are going, Am. J. Hypertens. 31 (12) (2018) 1247–1254, https://doi.org/10.1093/ajh/hpy148.

[24] Y. Zhao, Y. Hu, J.P. Smith, J. Strauss, G. Yang, Cohort profile: the China health and retirement longitudinal study (CHARLS), Int. J. Epidemiol. 43 (1) (2014) 61–68, https://doi.org/10.1093/ije/dys203.

[25] H. Chen, A.C. Mui, Factorial validity of the center for epidemiologic studies depression Scale short form in older population in China, Int. Psychogeriatr. 26 (1) (2014) 49–57, https://doi.org/10.1017/s1041610213001701.

[26] A.D. Association, 2. Classification and diagnosis of diabetes: standards of medical care in diabetes-2021, Diabetes Care 44 (Suppl 1) (2021) S15–s33, https://doi.org/10.2337/dc21-S002.

[27] Y. Yang, W. Piao, K. Huang, H. Fang, L. Ju, L. Zhao, D. Yu, Y. Ma, Dietary pattern associated with the risk of hyperuricemia in Chinese elderly: result from China nutrition and health surveillance 2015-2017, Nutrients 14 (4) (2022), https://doi.org/10.3390/nu14040844.

[28] D.H. Wolpert, W.G. Macready, No free lunch theorems for optimization, IEEE Trans. Evol. Comput. 1 (1) (1997) 67–82, https://doi.org/10.1109/4235.585893.

[29] P. Favaro, A. Vedaldi, AdaBoost, Computer Vision: A Reference Guide, Springer International Publishing, Cham, 2020, pp. 1–4.

[30] Y. Freund, R.E. Schapire, A decision-theoretic generalization of on-line learning and an application to boosting, JCSS 55 (1) (1997) 119–139, https://doi.org/10.1006/jcss.1997.1504.

[31] T. Bayes, An essay towards solving a problem in the doctrine of chances. 1763, MD Comput 8 (3) (1991) 157–171, https://doi.org/10.1016/B978-044450871-3/50096-6.

[32] J. Korstanje, Gradient Boosting with XGBoost and LightGBM, Advanced Forecasting with Python: with State-Of-The-Art-Models Including LSTMs, Facebook's Prophet, and Amazon's DeepAR, Apress, Berkeley, CA, 2021, pp. 193–205.

[33] S.J. Rigatti, Random forest, J. Insur. Med. 47 (1) (2017) 31–39, https://doi.org/10.17849/insm-47-01-31-39.1.

[34] J. Cervantes, F. García-Lamont, L. Rodríguez, A. Lopez-Chau, A comprehensive survey on support vector machine classification: applications, challenges and trends, Neurocomputing 408 (2020), https://doi.org/10.1016/j.neucom.2019.10.118.

[35] O. Kramer, K-Nearest Neighbors, Dimensionality Reduction with Unsupervised Nearest Neighbors, Springer Berlin Heidelberg, Berlin, Heidelberg, 2013, pp. 13–23.

[36] J.N. Mandrekar, Receiver operating characteristic curve in diagnostic test assessment, J. Thorac. Oncol. 5 (9) (2010) 1315–1316, https://doi.org/10.1097/JTO.0b013e3181ec173d.

[37] K. Goyal, S. Dumancic, H. Blockeel, Feature interactions in XGBoost. https://arxiv.org/pdf/2007.05758, 2020. (Accessed 30 April 2024).

[38] R. Dey, R. Mathur, Ensemble Learning Method Using Stacking with Base Learner, A Comparison, Springer Nature Singapore, Singapore, 2023, pp. 159–169, https://doi.org/10.1007/978-981-99-3878-0_14.

[39] X. Guo, L. Zou, X. Zhang, J. Li, L. Zheng, Z. Sun, J. Hu, N.D. Wong, Y. Sun, Prehypertension: a meta-analysis of the epidemiology, risk factors, and predictors of progression, Tex. Heart Inst. J. 38 (6) (2011) 643–652.

[40] W. Chen, Y. Guo, X. Yao, D. Zhao, Correlation of blood lipid and serum inflammatory factor levels with hypertensive disorder complicating pregnancy, Front Surg 9 (2022) 917458, https://doi.org/10.3389/fsurg.2022.917458.

[41] M. De Bhailis Á, P.A. Kalra, Hypertension and the kidneys, Br. J. Hosp. Med. 83 (5) (2022) 1–11, https://doi.org/10.12968/hmed.2021.0440.

[42] S.P. Kauhanen, T. Liimatainen, E. Kariniemi, M. Korhonen, J. Parkkonen, J. Vienonen, R. Vanninen, M. Hedman, A smaller heart-aorta-angle associates with ascending aortic dilatation and increases wall shear stress, Eur. Radiol. 30 (9) (2020) 5149–5157, https://doi.org/10.1007/s00330-020-06852-3.

[43] C. Liu, Z. Ye, L. Chen, H. Wang, B. Wu, D. Li, S. Pan, W. Qiu, H. Ye, Interaction effects between sleep-related disorders and depression on hypertension among adults: a cross-sectional study, BMC Psychiatr. 24 (1) (2024) 482, https://doi.org/10.1186/s12888-024-05931-9.

[44] W. Wu, W. Wang, Y. Gu, Y. Xie, X. Liu, X. Chen, Y. Zhang, X. Tan, Sleep quality, sleep duration, and their association with hypertension prevalence among low-income oldest-old in a rural area of China: a population-based study, J. Psychosom. Res. 127 (2019) 109848, https://doi.org/10.1016/j.jpsychores.2019.109848.

[45] F. Lopez-Jimenez, W. Almahmeed, H. Bays, A. Cuevas, E. Di Angelantonio, C.W. le Roux, N. Sattar, M.C. Sun, G. Wittert, F.J. Pinto, J.P.H. Wilding, Obesity and cardiovascular disease: mechanistic insights and management strategies. A joint position paper by the World Heart Federation and World Obesity Federation, Eur J Prev Cardiol 29 (17) (2022) 2218–2237, https://doi.org/10.1093/eurjpc/zwac187.

[46] A. Angi, F. Chiarelli, Obesity and diabetes: a sword of damocles for future generations, Biomedicines 8 (11) (2020), https://doi.org/10.3390/biomedicines8110478.

[47] Y. Li, J. Wang, L. Zhao, Z. Wang, D. Yu, Y. He, G. Ding, The drinking status and associated factors in adults in China, Chin. J. Epidemiol. (7) (2018) 898–903, https://doi.org/10.3760/cma.j.issn.0254-6450.2018.07.007.

[48] X. Shang, W. Wang, S. Keel, J. Wu, M. He, L. Zhang, Leading determinants for disease-free status in community-dwelling middle-aged men and women: a 9-year follow-up cohort study, Front. Public Health 7 (2019) 320, https://doi.org/10.3389/fpubh.2019.00320.

[49] D. Kocevska, T.S. Lysen, A. Dotinga, M.E. Koopman-Verhoeff, M. Luijk, N. Antypa, N.R. Biermasz, A. Blokstra, J. Brug, W.J. Burk, H.C. Comijs, E. Corpeleijn, H. S. Dashti, E.J. de Bruin, R. de Graaf, I.P.M. Derks, J.F. Dewald-Kaufmann, P.J.M. Elders, R. Gemke, L. Grievink, L. Hale, C.A. Hartman, C.J. Heijnen, M. Huisman, A. Huss, M.A. Ikram, S.E. Jones, M.K. Velderman, M. Koning, A.M. Meijer, K. Meijer, R. Noordam, A.J. Oldehinkel, J.O. Groeniger, B. Penninx, H.S.J. Picavet, S. Pieters, S.A. Reijneveld, E. Reitz, C.M. Renders, G. Rodenburg, F. Rutters, M.C. Smith, A.S. Singh, M.B. Snijder, K. Stronks, M. Ten Have, J.W.R. Twisk, D. Van de Mheen, J. van der Ende, K.B. van der Heijden, P.G. van der Velden, F.J. van Lenthe, R.R.L. van Litsenburg, S.H. van Oostrom, F.J. van Schalkwijk, C. M. Sheehan, R.A. Verheij, F.C. Verhulst, M.C.M. Vermeulen, R.C.H. Vermeulen, W.M.M. Verschuren, T.G.M. Vrijkotte, A.H. Wijga, A.M. Willemen, M. Ter Wolbeek, A.R. Wood, Y. Xerxa, W.M. Bramer, O.H. Franco, A.I. Luik, E.J.W. Van Someren, H. Tiemeier, Sleep characteristics across the lifespan in 1.1 million people from The Netherlands, United Kingdom and United States: a systematic review and meta-analysis, Nat Hum Behav 5 (1) (2021) 113–122, https://doi.org/10.1038/s41562-020-00965-x.

[50] Ş. Alpsoy, Exercise and hypertension, Adv. Exp. Med. Biol. 1228 (2020) 153–167, https://doi.org/10.1007/978-981-15-1792-1_10.

[51] H. Xiao, Y. Yan, Y. Gu, Y. Zhang, Strategy for sodium-salt substitution: on the relationship between hypertension and dietary intake of cations, Food Res. Int. 156 (2022) 110822, https://doi.org/10.1016/j.foodres.2021.110822.