Public Health 247 (2025) 105889

Contents lists available at ScienceDirect

Public Health

journal homepage: www.elsevier.com/locate/puhe

Original Research

Development of a machine learning model to predict the probability of health checkup participation in Japan

Asuka Oyama [a,b], Midori Noguchi [c,*]

[a] Health and Counseling Center, The University of Osaka, Osaka, Japan
[b] Epidemiology Section, Division of Public Health, Osaka Institute of Public Health, Osaka, Japan
[c] Division of Public Health, Department of Social Medicine, Graduate School of Medicine, The University of Osaka, Osaka, Japan

ARTICLE INFO

Keywords:
Specific health checkup
Checkup attendance
Machine learning
Health checkup participation rate

ABSTRACT

Objectives: Enhancing health checkup participation is crucial for early detection and treatment of non-communicable diseases and for improving public health. Effectively increasing health checkup rates requires identifying and encouraging individuals likely to adopt health-oriented behaviours. We aimed to develop a machine learning model to predict the participation probability in a specific health checkup in the following year.
Study design: Retrospective cohort study.
Methods: We analysed data from 58,863 National Health Insurance-insured individuals in Kochi Prefecture, Japan, who underwent specific health checkups during the fiscal years (FYs) 2013–2017. The dataset includes physical measurements, blood pressure measurements, blood and urine tests, and self-reported questionnaires. Predictive models for FY2018 participation were developed using LightGBM and evaluated using the area under the receiver operating characteristic curve (AUC) and reliability curves. SHAP was used to assess the feature's importance. External validation for FY2019 and FY2020 assessed temporal robustness.
Results: Predictive accuracy for FY2018 was high, with AUCs of 0.824 (95 % confidence interval [95 % CI]: 0.813–0.835) for men and 0.820 (95 % CI: 0.810–0.830) for women. External validation of FY2019 showed AUCs of 0.821 and 0.807 for men and women, respectively. In FY2020, prediction accuracy declined, with AUCs of 0.798 and 0.794 for men and women, respectively. Key predictive features included years since the last checkup, past checkup frequency, age, systolic blood pressure, and lifestyle factors.
Conclusions: By developing an accurate model to predict future health checkup participation, we identified a novel indicator that enables efficient, optimized recommendations and may help improve participation rates.

1. Introduction

Health checkups are routine examinations that assess overall health and aid in the early detection of potential health issues through physical and blood tests. Additionally, they promote healthier lifestyles and provide opportunities for treatment through counselling.[1–3] Although no consensus exists on the overall impact and necessity of health checkups,[2–4] regular health checkups assist in the early detection of diseases and the reduction of risk factors in many developed countries.[2]

In Japan, specific health checkups and guidance were initiated in the fiscal year (FY) 2008 for the screening and early detection of lifestyle-related diseases.[5,6] As Japan has a universal health insurance system, almost all citizens can access these specific health checkups and receive guidance at a low cost. The implementation of specific health guidance for individuals with obesity accompanied by hypertension, hyperglycaemia, dyslipidaemia, or smoking habits (as identified through health checkup results) has been reported to help prevent the progression of metabolic syndrome,[7,8] mitigate risk factors for cardiovascular disease (CVD)[9] and increase doctor visits.[10,11] Therefore, participation in specific health checkups is essential for the reliable screening of individuals at high risk for diabetes and CVDs.

Despite its importance, the participation rate for specific health checkups remains low. The target participation rate for these checkups, set by the Ministry of Health, Labor and Welfare, was ≥70 %. However,

* Corresponding author. Division of Public Health, Department of Social Medicine, Graduate School of Medicine, The University of Osaka, 2-2, Yamadaoka, Suita-shi, Osaka, 565–0871, Japan.
E-mail address: noguchi@pbhel.med.osaka-u.ac.jp (M. Noguchi).

https://doi.org/10.1016/j.puhe.2025.105889
Received 6 June 2025; Received in revised form 14 July 2025; Accepted 22 July 2025
Available online 7 August 2025
0033-3506/

the participation rate for specific health checkups in FY2022 was only 57.8 %.[12] Undergoing a specific health checkup is not only a means of detecting and preventing progression of lifestyle-related diseases at an early stage but also a valuable opportunity to promote behavioural modification through targeted health guidance for individuals identified as at risk for conditions including prediabetes and subclinical cardiovascular disease. These checkups serve as a critical foundation for proactive health management at the individual level and increasing participation constitutes an urgent priority in public health policy.

Although various strategies, such as reminder letters and phone calls implemented with unique approaches by local municipalities, have been explored to increase participation in specific health checkups, no consistently effective method has yet been established.[13,14] One contributing factor is that participation in health checkups is influenced by various factors, such as age, sex, health condition, and socioeconomic status.[15–17] Comprehensive consideration of these factors and accurate identification of individuals who would benefit from policies aimed at increasing health checkup rates is challenging. Therefore, in this study, we used a machine learning model to calculate the probability that an individual would attend a specific health checkup. Multiple factors can be accounted for by assigning probabilities, and policies based on these probabilities can possibly achieve maximum impact. Although previous analyses have predicted health checkup attendance and hospital visits,[18–20] the present study is the first to predict individual-level probabilities using large-scale real-world data. Additionally, we identified key factors influencing predictions and examined their relative importance.

2. Methods

2.1. Participants

This study aimed to predict the probability of undergoing a health checkup in the following year based on a 5-year history of specific health checkups and their items. Study participants were individuals enrolled in the National Health Insurance (NHI) program in Kochi Prefecture, Japan. We selected insured individuals from this group who underwent a specific health checkup at least once in the 5 years from FY2013 to FY2017. We included those enrolled in the NHI in FY2018 with consistent birth dates and sex records. Additionally, we excluded individuals aged ≥75 years who were transferred to the late-stage elderly medical care system at the end of FY2018 and those <45 years old without an available 5-year participation history.

2.2. Predictors

All individuals covered by the NHI are eligible for a specific health checkup once a year. Health checkup data included over 50 items, such as physical measurements, blood pressure, blood tests, urine tests, and responses to medical questionnaires. First, if an individual had received multiple specific health checkups in the past 5 years, only the most recent health checkup data were used. From these data, we constructed a dataset for analysis to predict the probability of health checkups in the following year. Second, we excluded variables with missing data rates >50 % (e.g., fasting plasma glucose, haematocrit, haemoglobin, and red blood cell count). Next, we calculated the correlation coefficients among variables and removed variables highly correlated with others (e.g., height, weight, waist circumference, diastolic blood pressure, aspartate aminotransferase, gamma-glutamyl transferase, serum creatinine, metabolic syndrome, and alcohol consumption). Subsequently, we calculated the total number of checkups and the time elapsed since the last checkup over the past 5 years based on the checkup history. Finally, the health checkup items used for the prediction included age, sex, body mass index (BMI), systolic blood pressure (SBP), triglycerides (TG), high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C), alanine aminotransferase (ALT), haemoglobin A1c (HbA1c), uric acid (UA), estimated glomerular filtration rate (eGFR), urinary glucose (UGLU), urinary protein (UPRO), and health guidance level (Guidance Level), such as "Motivational support (single health guidance session) and "Active support" (follow-up through 3-month health guidance). Additional variables included anti-diabetes mellitus (anti-DM) drugs, anti-hypertension (anti-HTN) drugs, anti-dyslipidaemia (anti-DLP) drugs, and a medical history of stroke, heart attack, and chronic kidney disease. Lifestyle-related factors also included smoking, weight gain since age 20 years (weight gain), weight change in the past year (weight change), regular exercise habits, daily physical activity, walking speed compared with the same age group, fast-eating habits, late-night dinner habits, late-night snacking, skipping breakfast, alcohol consumption frequency, restorative sleep quality, lifestyle improvement intention, willingness to receive health guidance, checkups in the past 5 years, and the timing of the last checkup. Details of the questionnaires are provided in Supplementary File 1.

As a preprocessing step, values outside the 0.01st and 99.99th percentiles of the continuous variables were treated as missing.

2.3. Prediction model

In this study, we used LightGBM to predict the probability of participation in a specific health checkup in FY2018.[21,22] This nonlinear decision-tree-based model is known to outperform logistic regression by increasing the model's degrees of freedom.[23] LightGBM has the advantages of being less dependent on the type and number of explanatory variables, handling missing data, and being less susceptible to outliers.

The LightGBM Package in Python was used to implement prediction models. The hyperparameters were optimized using Optuna software.[24] To avoid overfitting, early stopping was applied when the loss of the validation set did not improve for over 50 consecutive iterations during the training process. Hyperparameter optimization was conducted using stratified k-fold cross-validation with $k = 5$.

2.4. Validation and evaluation

We created separate prediction models for men and women to investigate the differences in factors affecting participants' probability of undergoing health checkups by sex. We used the reliability curve and area under the receiver operating characteristic (ROC) curve (AUC) to assess model performance. A reliability curve is an indicator that visually estimates the calibration of predicted probabilities. It plots the predicted probability on the horizontal axis and the actual participation rate on the vertical axis, visually representing how well the predicted probabilities of a model match the observed outcomes. The ROC curve was used to assess the performance of classification ability. It plots the true-positive rate against the false-positive rate at various thresholds. In addition, the AUC measures the classification performance, with values closer to one indicating higher discrimination performance.

The analysis data were randomly split into 80 % training and 20 % test data to verify prediction accuracy. The proportion of individuals who underwent health checkups was stratified to equal the overall proportion in the dataset. The model was trained using the training data, and its performance was evaluated using the test data.

2.5. Feature importance

In many AI tasks, increasing model complexity often makes the interpretation of results challenging. Therefore, we used Shapley additive explanations (SHAP) to quantify each variable.[25,26] SHAP is a method that distributes the predicted probability to each variable based on its contribution. This analysis used the SHAP to quantify how each explanatory variable influenced the predicted probability of undergoing a health checkup in the following year.

2.6. Statistical analysis

Continuous variables were reported as the mean ± standard deviation (SD) for normally distributed variables and the median and interquartile range (IQR) for non-normally distributed variables. Categorical variables were reported as counts and proportions. Due to the large sample size, the normality of the continuous variables was assessed visually using histograms and Q-Q plots rather than statistical tests. Based on this visual assessment, age, TG, ALT, and HbA1c were treated as non-normally distributed variables. The histograms and Q-Q plots for each variable are presented in Supplementary Files 2 and 3.

We constructed a prediction model for the health checkup participation the following year for each sex using LightGBM. Missing values were retained during the training because LightGBM can handle missing values without imputation. We evaluated predictive performance using the AUC and reliability curves. For the reliability curves, the predicted probabilities were divided into 10 equal-width bins ranging from 0 to 1, and the calibration ability was assessed within each bin. We considered a prediction statistically significant if the 95 % confidence interval (CI) of each evaluation point crossed the 45-degree diagonal line.

SHAP values were calculated for all test data to assess the overall importance of each variable. To examine age-related differences, we then divided the test data into two age groups (≥65 years and <65 years) and compared variable importance using SHAP values.

For external validation, the model developed in this study was applied to predict health checkup participation rates for FY2019 and FY2020, and its robustness was assessed. All analyses were performed using Python version 3.10.

3. Results

3.1. Participant characteristics

This study included 89,695 local NHI-insured participants in Japan who underwent a specific health checkup between FY2013 and FY2017. Of these, 58,863 continuously enrolled participants during FY2018 were included in the analysis.

The baseline characteristics of the 58,863 participants are presented in Table 1. Among the participants, 26,235 (44.6 %) and 32,628 (55.4 %) were men and women, respectively, with a mean age of 66 years (IQR: 61–69). Of these, 40,436 (69.3 %) underwent a specific health checkup in FY2018, whereas 18,437 (30.7 %) did not. The missing rates for the variables ranged from 0.0 % to 0.4 %, with the highest rate observed for skipping breakfast. Other baseline characteristics, such as questionnaires and history of specific health checkups, are presented in Supplementary File 4.

3.2. Model performance

The model performance was evaluated based on the AUC and reliability curves. The AUC was 0.824 (95 %CI: 0.813–0.835) for men and 0.820 (0.810–0.830) for women. Fig. 1 illustrates the reliability curves for men and women. As shown in Fig. 1, the 95 % CIs for all evaluation points intersected the 45° line.

3.3. Feature importance

Fig. 2 shows the importance of the variables calculated using SHAP. This highlights that the number of years since the most recent participation is the most important feature for predicting the probability of health checkup participation for both men and women. The probability tended to decrease over time as the time since the last participation lengthened. Next, the number of times an individual participated in the previous 5 years was another major contributor to the predicted probability, with more frequent participation in the past associated with higher probabilities. Subsequently, older age and lower SBP also contributed to a higher predicted participation probability. Supplementary File 5 presents the results stratified according to whether the participants were aged ≥65 years. Although no substantial differences in the top-ranking variables were observed, SBP had a higher feature importance than did age in the group aged ≥65 years.

Fig. 3a and b shows the relationship between the time elapsed since the last checkup and the probability of receiving a checkup in the following year. Of those who underwent a checkup 1 year prior, approximately 80 % underwent a checkup the following year. After 2 years, the rate of checkups fell to 40–50 %; after ≥4 years, the probability of participation in the following year dropped to approximately 20 %. Similarly, Fig. 3c and d plot the relationship between the number of health checkups received in the past 5 years and the probability of a checkup in the following year. The rate of receiving health checkups increased with the frequency of previous health checkups. The rate was approximately 40 % for those with only one past checkup, whereas those with five past checkups had a rate of >90 %.

3.4. External validation

In this analysis, we used specific health checkup data from FY2013 to FY2017 to predict the participation rate in FY2018. As an external validation, we assessed the model's performance in predicting the participation probability for FY2019 and FY2020 using a model trained on data from FY2013 to FY2017. The AUC for FY2019 was 0.821 (95 % CI: 0.816–0.826) for men and 0.807 (95 % CI: 0.802–0.811) for women. Similarly, the AUC for FY2020 was 0.798 (95 % CI: 0.793–0.803) for men and 0.794 (95 % CI: 0.789–0.798) for women. Supplementary File 6 shows the reliability curves for specific health checkup probabilities for FY2019 and FY2020. In FY2019, as in FY2018, the 95 % CIs intersected the 45° line for many evaluation points, whereas in FY2020, the

Table 1
Baseline characteristics include body measurements, blood pressure tests, blood tests, urine tests, and health guidance levels.

	level	Non-attendance	Attendance
n		18,427	40,436
Age, median [IQR]		65 [58, 68]	67 [63, 69]
Sex, n (%)	Men	8792 (33.5)	17,443 (66.5)
	Women	9635 (29.5)	22,993 (70.5)
BMI, mean (SD)		23.6 (3.8)	23.2 (3.5)
SBP, mean (SD)		131.4 (19.5)	129.0 (18.0)
TG, median [IQR]		112.0 [78.0, 167.0]	106.0 [75.0, 153.0]
HDL-C, mean (SD)		61.0 (16.5)	63.0 (16.4)
LDL-C, mean (SD)		121.1 (33.1)	120.3 (30.0)
ALT, median [IQR]		19.0 [14.0, 26.0]	18.0 [14.0, 25.0]
HbA1c, median [IQR]		5.7 [5.4, 6.0]	5.7 [5.4, 6.0]
UA, mean (SD)		5.3 (1.4)	5.3 (1.3)
eGFR, mean (SD)		74.7 (17.2)	72.3 (15.0)
UGLU, n (%)	–	17,073 (30.8)	38,394 (69.2)
	±	257 (36.1)	454 (63.9)
	1+	306 (37.6)	508 (62.4)
	2+	298 (41.7)	416 (58.3)
	3+	411 (40.9)	595 (59.1)
UPRO, n (%)	–	15,648 (30.6)	35,560 (69.4)
	±	1597 (33.1)	3223 (66.9)
	1+	688 (38.8)	1085 (61.2)
	2+	315 (45.1)	384 (54.9)
	3+	98 (45.8)	116 (54.2)
Guidance Level, n (%)	None	14,777 (30.0)	3,4481 (70.0)
	Motivational	2285 (34.0)	4428 (66.0)
	Active	1346 (47.1)	1513 (52.9)

ALT, alanine aminotransferase; BMI, body mass index; eGFR, estimated glomerular filtration rate; HbA1c, haemoglobin A1c; HDL-C, high-density lipoprotein cholesterol; IQR, interquartile range; LDL-C, low-density lipoprotein cholesterol; SBP, systolic blood pressure; SD, standard deviation; TG, triglycerides; UA, uric acid; UGLU, urine glucose; UPRO, urine protein.

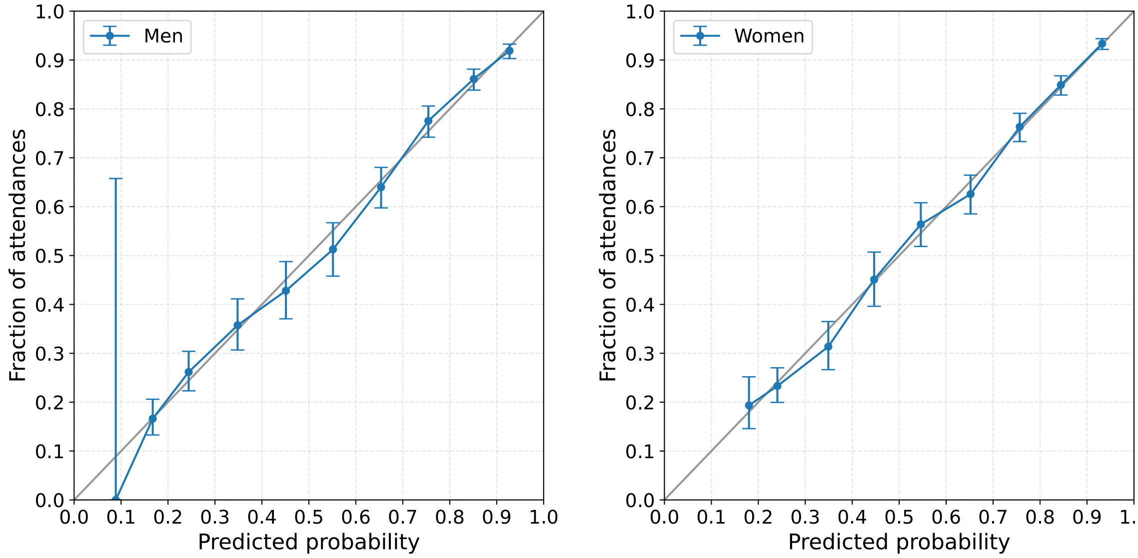

Fig. 1. Reliability curves for men (left) and women (right). The horizontal axis is the predicted probabilities. The vertical axis is a fraction of attendance. The error bars represent the 95 % confidence intervals.

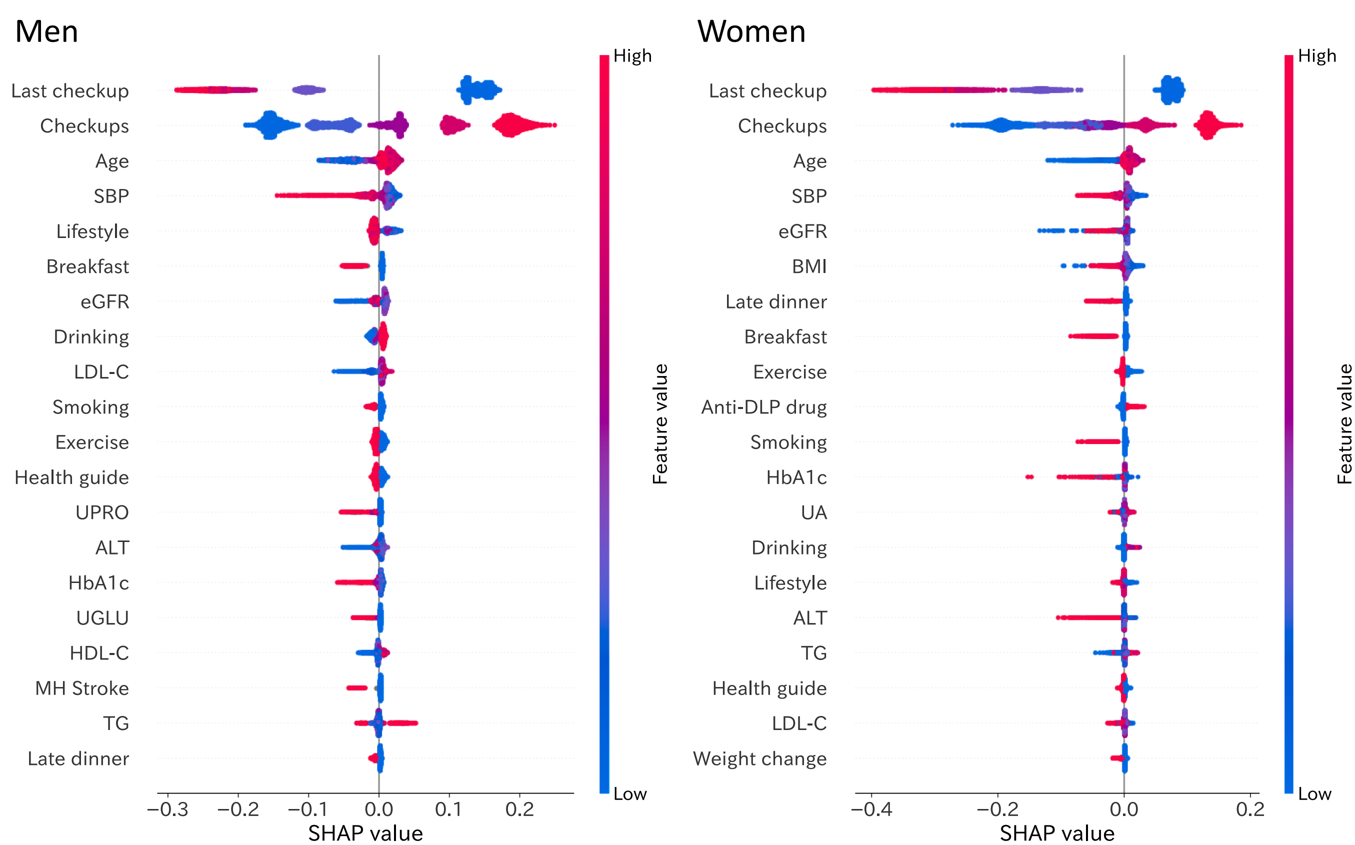

Fig. 2. Feature importance for men (left) and women (right). The top 20 most important variables are shown. The x-axis represents the SHAP values, and the y-axis lists the variables in descending order of their overall contribution. For each variable, a red dot indicates that the individual's value is relatively high, while a blue dot indicates that the value is relatively low. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

probability of receiving a specific health checkup tended to be overestimated.

4. Discussion

Herein, we predicted the probability of receiving a specific health checkup in the following year using health checkup data and a 5-year history. Internal validation revealed minimal differences between men

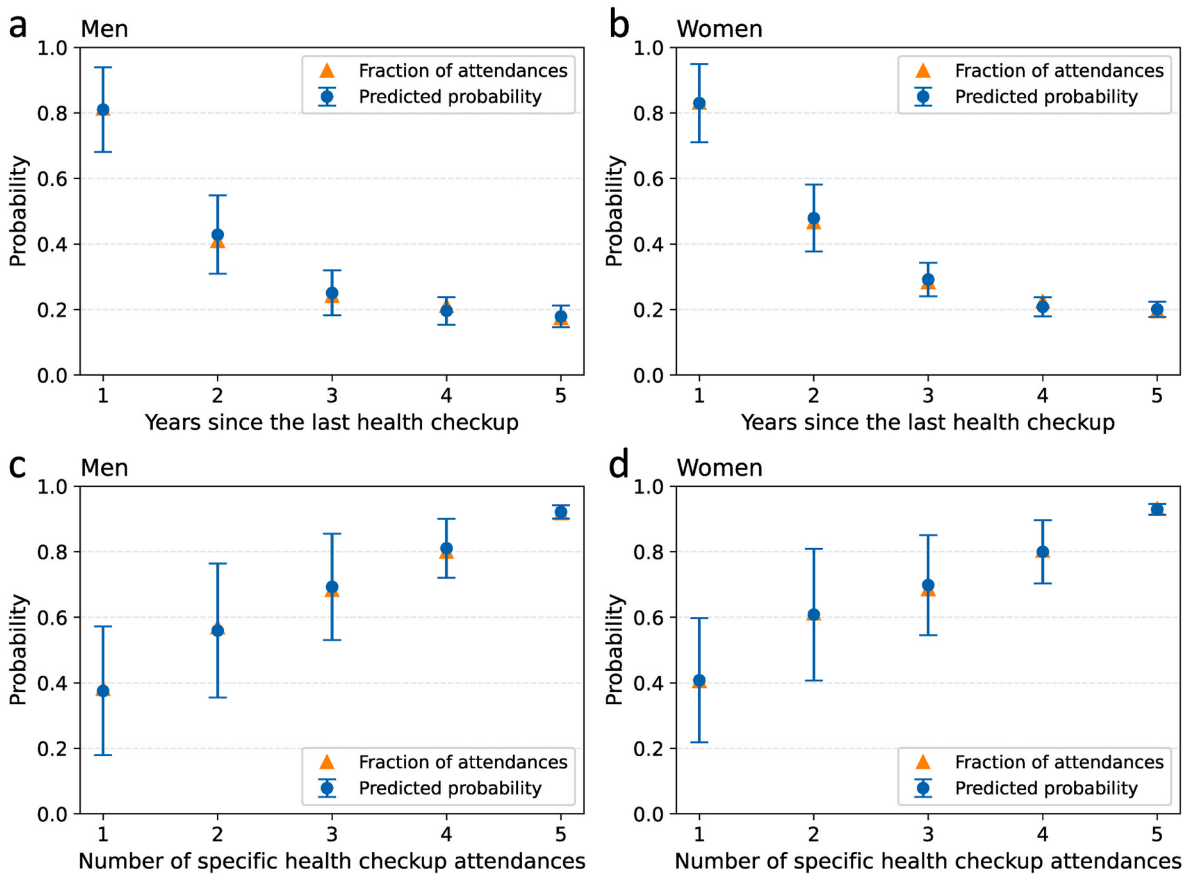

Fig. 3. Past specific health checkup attendance and predicted probability for men (left) and women (right), shown with mean ± standard deviation. Years since the last health checkup (top) and the number of specific health checkup attendances (bottom).

and women, demonstrating high discrimination in both groups. Moreover, the reliability curve showed that the 95 % CIs intersected the 45° line across all evaluation points, confirming high predictive accuracy across all probability groups.

The SHAP values showed that the timing of the last health checkup had the strongest association with future participation. Additionally, the probability of participating in health checkups increases with age. These findings suggest that health beliefs, health concerns, and routine participation in health checkups are related to the number of previous health checkups, which is a key factor in future participation in health checkups. Focusing on SBP, the probability of health checkups increased in groups with lower blood pressure for both men and women. This may be because individuals with high blood pressure have already started treatment or, if untreated, are concerned about receiving strict health advice from doctors and public health nurses. A similar trend was observed for BMI. However, the importance of BMI was lower in men, whereas it was higher in women. Previous studies have shown that women are more likely to perceive themselves as overweight than men.[27–29] This suggests that concerns about obesity-related health advice are more pronounced in women than in men.

Next, we examined items in the questionnaire related to daily life. The study revealed that the probability of undergoing a specific health checkup is higher among groups with healthy lifestyles, such as those who habitually eat breakfast, do not smoke, and exercise regularly. This suggests that groups with a lower probability of undergoing health checkups tend to have poorer lifestyle habits, and a low rate of health checkups is strongly linked to a lack of concern for one's health.

The highly accurate prediction of health checkup rates is expected to encourage participation. Tailoring strategies for promoting health checkups and optimizing recommendation frequencies for groups with varying probabilities may efficiently maximize uptake under limited cost and time constraints. For instance, groups with high participation probability may not require additional recommendations, whereas those with low or moderate rates could benefit from targeted interventions and nudging techniques, such as personalized reminders, to enhance participation.

In previous studies, the main factors affecting whether people underwent health checkups included male sex, younger age, low socioeconomic status, and low interest in health checkups.[16,17] Consistent with these findings, our study also found that younger individuals were less likely to undergo specific health checkups. Furthermore, the number of past checkups and the time since the last checkup—both indicators of health beliefs—were associated with checkup participation, providing additional insight into behavioural determinants of health checkup attendance. Moreover, self-reported lifestyle habits indicated that groups with poorer lifestyles were less likely to undergo health checkups, which is consistent with previous findings.[15–17] This study found that blood pressure levels were strongly associated with the probability of receiving specific health checkups. Given previous reports showing that the onset of hypertension, a cardiovascular risk factor, is associated with lower participation in health checkups,[17] targeted strategies to encourage checkup attendance among individuals with high blood pressure may be necessary.

In contrast, this study found that both men and women who regularly drink alcohol were slightly more likely to undergo health checkups. Previous studies suggested that higher alcohol consumption reduces health checkup rates.[16] However, these results must be interpreted with caution. The alcohol-related items in this analysis referred to frequency rather than consumption. Therefore, frequent drinking does not necessarily indicate high alcohol intake. Although the questionnaire included questions on alcohol consumption, due to high missing values and multicollinearity, alcohol consumption was excluded from the analysis. Future studies should incorporate quantitative measures.

Herein, we developed a model to predict health checkup probability for FY2018 and confirmed its reliability via internal validation. We then evaluated the model's temporal robustness by applying it to FY2019 and FY2020 through external validation. In FY2019, performance was comparable to FY2018. However, in FY2020, the model showed a trend of overestimation. This trend is likely attributable to the global spread of coronavirus disease. In Japan, a state of emergency was imposed in major cities including in Kochi Prefecture, many residents refrained from going out, reducing health checkup participation in FY2020. When using this model to predict health checkup probabilities for FY2021 or later, the predicted probabilities may be underestimated, as using FY2020 data for prediction reflects reduced participation due to the pandemic. Therefore, rebuilding the model to account for the conditions observed in FY2020 is necessary for accurate predictions after FY2021.

The strength of this study is its large sample size. However, following limitations are worth mention: first, data were obtained from a non-urban area, limiting generalizability to other regions of Japan, particularly urban areas with differing demographic and socioeconomic characteristics; external validity should be assessed in future research; second, the analysis did not consider whether participants were receiving medical treatment, as those under treatment typically do not attend specific health checkups; and third, individuals who had not undergone a specific health checkup in the past 5 years were excluded, so the calculated annual health checkup uptake rate applies only to those who have been checked at least once in that period.

In this study, we accurately predicted health checkup probabilities and computed variable importance using a machine learning model and an explanatory technique. The time elapsed since the last checkup and the number of previous checkups were the most influential factors in increasing participation likelihood. By estimating future participation probabilities, we developed a novel indicator that enables more efficient and optimized health checkup recommendations.

Author statements

Acknowledgments

We gratefully acknowledge the cooperation of Kochi Prefecture in providing access to the KDB data that formed the basis of this research.

Author contributions

A.O. and M.N. conceived and designed the study. A.O. and M.N. wrote the manuscript. A.O. and M.N. designed the data analysis framework. A.O. performed data cleaning and all the analyses. All authors contributed to the discussion and reviewed the manuscript.

Ethical approval

The study protocol was approved by the Ethics Committee of the University of Osaka Hospital Ethical Review Board (20209(T1)).

Data statement

Data are available from Kochi Prefecture upon reasonable request and with approval from the corresponding author and the data provider.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Competing interests

The authors declare no conflicts of interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.puhe.2025.105889.

References

1. Virgini V, Meindl-Fridez C, Battegay E, Zimmerli LU. Check-up examination: recommendations in adults. *Swiss Med Wkly*. 2015;145, w14075. https://doi.org/10.4414/smw.2015.14075.
2. Liss DT, Uchida T, Wilkes CL, Radakrishnan A, Linder JA. General health checks in adult primary care: a review. *JAMA*. 2021;325:2294–2306. https://doi.org/10.1001/jama.2021.6524.
3. Krogsbøll LT, Jørgensen KJ, Gøtzsche PC. General health checks in adults for reducing morbidity and mortality from disease. *JAMA*. 2013;309:2489–2490. https://doi.org/10.1001/jama.2013.5039.
4. Higa S, Ii Y, Nozawa K, Yamamoto Y, Ohwaki K, Asami Y. Relationship of annual health checkups with the subject's subsequent behavior of cardiovascular risk management in a real-world setting in Japan: a retrospective cohort study on changes in antihypertensive drug prescription and blood pressure from 2015 to 2017. *Drugs Real World Outcomes*. 2021;8:215–225. https://doi.org/10.1007/s40801-020-00224-5.
5. Yamagishi K, Iso H. The criteria for metabolic syndrome and the national health screening and education system in Japan. *Epidemiol Health*. 2017;39, e2017003. https://doi.org/10.4178/epih.e2017003.
6. Kohro T, Furui Y, Mitsutake N, et al. The Japanese National health screening and intervention program aimed at preventing worsening of the metabolic syndrome. *Int Heart J*. 2008;49:193–203. https://doi.org/10.1536/ihj.49.193.
7. Takeuchi Y, Kashiwabara K, Hosoi H, Imai H, Matsuyama Y. Longitudinal effects of a nationwide lifestyle intervention program on cardiometabolic outcomes in Japan: an observational cohort study. *Prev Med*. 2020;141, 106301. https://doi.org/10.1016/j.ypmed.2020.106301.
8. Tsushita K, S Hosler A, Miura K, et al. Rationale and descriptive analysis of specific health guidance: the nationwide lifestyle intervention program targeting metabolic syndrome in Japan. *J Atherosclerosis Thromb*. 2018;25:308–322. https://doi.org/10.5551/jat.42010.
9. Noguchi M. Evaluation of public health nurse-led health counseling programs for the prevention of cardiovascular diseases in amagasaki City, Japan. *Public Health Nurs*. 2022;39:1213–1219. https://doi.org/10.1111/phn.13110.
10. Iso H, Noguchi M, Yokoyama T, et al. Effect of a community-based program to accelerate referral to physicians for individuals at high risk of lifestyle-related diseases: a cluster randomized trial. *J Atherosclerosis Thromb*. 2023;30:1389–1406. https://doi.org/10.5551/JAT.64100.
11. Noguchi M, Kojima S, Sairenchi T, et al. Japan trial in high-risk individuals to enhance their referral to physicians (J-HARP)-a nurse-led, community-based prevention program of lifestyle-related disease. *J Epidemiol*. 2020;30:194–199. https://doi.org/10.2188/jea.JE20180194.
12. Ministry of Health, labour and welfare. https://www.mhlw.go.jp/stf/newpage_03092.html.
13. Bunten A, Porter L, Gold N, Bogle V. A systematic review of factors influencing NHS health check uptake: invitation methods, patient characteristics, and the impact of interventions. *BMC Public Health*. 2020;20:93. https://doi.org/10.1186/s12889-019-7889-4.
14. Kuwabara Y. Development and implementation of a program to promote regular specific health checkups for those who do not regularly receive checkups. *Health Behav Sci*. 2021;20:1–12.
15. de Waard AM, Wändell PE, Holzmann MJ, et al, SPIMEU Research Group. Barriers and facilitators to participation in a health check for cardiometabolic diseases in primary care: a systematic review. *Eur J Prev Cardiol*. 2018;25:1326–1340. https://doi.org/10.1177/2047487318780751.
16. Hoebel J, Starker A, Jordan S, Richter M, Lampert T. Determinants of health check attendance in adults: findings from the cross-sectional German Health Update (GEDA) study. *BMC Public Health*. 2014;14:913. https://doi.org/10.1186/1471-2458-14-913.
17. Dryden R, Williams B, McCowan C, Themessl-Huber M. What do we know about who does and does not attend general health checks? Findings from a narrative scoping review. *BMC Public Health*. 2012;12:723. https://doi.org/10.1186/1471-2458-12-723.
18. Breeze F, Hossain RR, Mayo M, McKelvie J. Predicting ophthalmic clinic non-attendance using machine learning: development and validation of models using nationwide data. *Clin Exp Ophthalmol*. 2023;51:764–774. https://doi.org/10.1111/ceo.14310.
19. Dashtban M, Li W. Predicting non-attendance in hospital outpatient appointments using deep learning approach. *Health Syst (Basingstoke)*. 2022;11:189–210. https://doi.org/10.1080/20476965.2021.1924085.
20. Shimoda A, Ichikawa D, Oyama H. Using machine-learning approaches to predict non-participation in a nationwide general health check-up scheme. *Comput Methods Progr Biomed*. 2018;163:39–46. https://doi.org/10.1016/j.cmpb.2018.05.032.

21. Ke G, Meng Q, Finley T, et al. LightGBM: a highly efficient gradient boosting decision tree. *Adv Neural Inf Process Syst.* 2017;30:52. Available from: https://github.com/Microsoft/LightGBM.
22. Friedman JH. Greedy function approximation: a gradient boosting machine. *Ann Stat.* 2001;29:1189–1232. https://doi.org/10.1214/aos/1013203451.
23. Seto H, Oyama A, Kitora S, et al. Gradient boosting decision tree becomes more reliable than logistic regression in predicting probability for diabetes with big data. *Sci Rep.* 2022;12, 15889. https://doi.org/10.1038/s41598-022-20149-z.
24. Akiba T, Sano S, Yanase T, Ohta T, Koyama M. Optuna: a next-generation hyperparameter optimization framework. In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining.* KDD'19; 2019: 2623–2631. https://doi.org/10.1145/3292500.3330701.
25. Lundberg S, Lee SI. A unified approach to interpreting model predictions. *Adv Neural Inf Process Syst.* 2017;2:4766–4775. http://arxiv.org/abs/1705.07874:2017-Decem.
26. Lundberg SM, Erion G, Chen H, et al. From local explanations to global understanding with explainable AI for trees. *Nat Mach Intell.* 2020;2:56–67. https://doi.org/10.1038/s42256-019-0138-9.
27. Inoue M, Toyokawa S, Miyoshi Y, et al. Degree of agreement between weight perception and body mass index of Japanese workers: my health up study. *J Occup Health.* 2007;49:376–381. https://doi.org/10.1539/joh.49.376.
28. Inoue M, Toyokawa S, Inoue K, et al. Lifestyle, weight perception and change in body mass index of Japanese workers: my health up study. *Public Health.* 2010;124: 530–537. https://doi.org/10.1016/j.puhe.2010.04.009.
29. Kagawa M, Kuroiwa C, Uenishi K, et al. A comparison of body perceptions in relation to measured body composition in young Japanese males and females. *Body Image.* 2007;4:372–380. https://doi.org/10.1016/j.bodyim.2007.06.002.