





Non-laboratory-based risk assessment model for case detection of diabetes mellitus and pre-diabetes in primary care

Weinan Dong¹, Tsui Yee Emily Tse^{1,2*} , Lynn Ivy Mak¹, Carlos King Ho Wong^{1,3} , Yuk Fai Eric Wan^{1,3}, Ho Man Eric Tang¹ , Weng Yee Chin¹, Laura Elizabeth Bedford¹, Yee Tak Esther Yu^{1,2}, Wai Kit Welchie Ko⁴, Vai Kiong David Chao⁵, Choon Beng Kathryn Tan⁶ , Lo Kuen Cindy Lam^{1,2}

¹Department of Family Medicine and Primary Care, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China, ²Department of Family Medicine, The University of Hong Kong Shenzhen Hospital, Shenzhen, China, ³Department of Pharmacology and Pharmacy, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China, ⁴Department of Family Medicine and Primary Healthcare, Hong Kong West Cluster, Hospital Authority, Hong Kong, China, ⁵Department of Family Medicine & Primary Health Care, United Christian Hospital & Tseung Kwan O Hospital, Hospital Authority, Hong Kong, China, and ⁶Department of Medicine, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China

Keywords

Case detection, Machine learning, Risk model

*Correspondence

Tsui Yee Emily Tse
Tel.: +852-2814-7484
Fax: +852-2814-7475
E-mail address:
emilyht@hku.hk

J Diabetes Investig 2022; 13: 1374–1386

doi: 10.1111/jdi.13790

ABSTRACT

Introduction: More than half of diabetes mellitus (DM) and pre-diabetes (pre-DM) cases remain undiagnosed, while existing risk assessment models are limited by focusing on diabetes mellitus only (omitting pre-DM) and often lack lifestyle factors such as sleep. This study aimed to develop a non-laboratory risk assessment model to detect undiagnosed diabetes mellitus and pre-diabetes mellitus in Chinese adults.

Methods: Based on a population-representative dataset, 1,857 participants aged 18–84 years without self-reported diabetes mellitus, pre-diabetes mellitus, and other major chronic diseases were included. The outcome was defined as a newly detected diabetes mellitus or pre-diabetes by a blood test. The risk models were developed using logistic regression (LR) and interpretable machine learning (ML) methods. Models were validated using area under the receiver-operating characteristic curve (AUC-ROC), precision-recall curve (AUC-PR), and calibration plots. Two existing diabetes mellitus risk models were included for comparison.

Results: The prevalence of newly diagnosed diabetes mellitus and pre-diabetes mellitus was 15.08%. In addition to known risk factors (age, BMI, WHR, SBP, waist circumference, and smoking status), we found that sleep duration, and vigorous recreational activity time were also significant risk factors of diabetes mellitus and pre-diabetes mellitus. Both LR (AUC-ROC = 0.812, AUC-PR = 0.448) and ML models (AUC-ROC = 0.822, AUC-PR = 0.496) performed well in the validation sample with the ML model showing better discrimination and calibration. The performance of the models was better than the two existing models.

Conclusions: Sleep duration and vigorous recreational activity time are modifiable risk factors of diabetes mellitus and pre-diabetes in Chinese adults. Non-laboratory-based risk assessment models that incorporate these lifestyle factors can enhance case detection of diabetes mellitus and pre-diabetes.

BACKGROUND

Diabetes mellitus (DM) is a major public health burden as it is common and chronic, and its complications including cardiovascular diseases, renal disease, and retinopathy can lead

to disabilities and premature mortality¹. Diabetes mellitus develops slowly and the progression from normal blood glucose to diabetes mellitus may take up to a decade². Pre-diabetes mellitus (pre-DM) refers to the condition where blood glucose is between normal and diabetic levels. Globally, the prevalence of diabetes mellitus was estimated to be 9.3%

Received 8 July 2021; revised 3 March 2022; accepted 10 March 2022

in 2019³, and the estimated prevalence of pre-diabetes mellitus was much higher, at 35% in American adults⁴ and 35.7% in Chinese adults⁵. More than 80% of people with pre-diabetes mellitus⁶ and over half with diabetes mellitus³ remain undiagnosed.

Pre-diabetes mellitus is important because it is a high-risk state for diabetes mellitus with an annual conversion rate of 5–10%, and an eventual conversion rate of 70%^{7,8}, and the hyperglycemia of pre-diabetes mellitus may damage the kidneys⁴ and blood vessels⁹ before the onset of diabetes mellitus. Early detection of pre-diabetes mellitus and the timely introduction of lifestyle interventions can prevent or delay the onset of diabetes mellitus and related complications⁷.

Screening of diabetes mellitus and pre-diabetes mellitus in the general population is not cost-effective¹⁰. The World Health Organization (WHO) recommends targeted opportunistic screening of diabetes mellitus in high-risk individuals during routine care¹¹. The Hong Kong Reference Framework for Diabetes Care for Adults in Primary Care Settings¹² adopts the American Diabetes Association (ADA) recommendation to screen for diabetes mellitus based on age, BMI, and the presence of any co-existing risk factors, which might not be cost-effective.

Several non-laboratory-based risk assessment models for diabetes mellitus have been developed and incorporated into diabetes mellitus prevention programs worldwide to improve the effectiveness and efficiency of case detection of high-risk individuals for further blood tests. The most widely used are the ADA Risk Test¹³, the Leicester Self-Assessment score adopted by the UK National Institute for Health and Care Excellence (NICE)¹⁴, the Australian type 2 diabetes risk assessment tool (AUSDRISK)¹⁵, and the Canadian Diabetes Risk Questionnaire (CANRISK)¹⁶, but these models developed based on Caucasian populations may not be applicable to the Chinese population¹⁷. The New Chinese Diabetic Risk Score (NCDRS)¹⁸ and the Non-invasive Diabetes Score (NDS)¹⁹ were developed from cohorts of Chinese adults and appeared to be more accurate than the ADA Risk Test for Chinese. However, these existing models are all intended for the risk assessment of diabetes mellitus and none has been developed for identifying pre-diabetes mellitus. These models include broadly similar risk factors such as age, sex, body mass index (BMI), blood pressure, and a few included lifestyles factors (i.e. physical activity, fruit and vegetable consumption)¹⁵. Recent studies have found other lifestyle factors, such as alcohol consumption²⁰ and sleep²¹ are associated with the risk of diabetes mellitus, but their contribution to risk assessment for diabetes mellitus and pre-diabetes mellitus have not been evaluated.

This study aimed to develop a non-laboratory-based risk assessment model that includes traditional risk factors and lifestyle factors for the detection of undiagnosed diabetes mellitus and pre-diabetes mellitus in Chinese adults in primary care.

METHODS

Study design and subjects

This was a cross-sectional study using data from the Hong Kong Population Health Survey (PHS) 2014/15 which was conducted by the Department of Health, HKSAR Government²². The PHS adopted a systematic replicated sampling method to recruit a representative sample of 12,022 people aged 15 or above from the Hong Kong general population. Each participant completed a face-to-face questionnaire survey consisting of questions on socio-demographics, self-reported health status, and lifestyle factors. Of these, 2,347 adults aged 15–84 were randomly selected to undergo physical measurements including blood pressure, weight, height, waist and hip circumference, and a blood test that included fasting plasma glucose and hemoglobin A1c (HbA1c). Of the 2,347 participants, we included 1,857 subjects without any self-reported doctor-diagnosis of diabetes mellitus or pre-diabetes mellitus, hypertension, cardiovascular diseases (CVD) (coronary heart disease, stroke), cancer, renal disease, or anemia in this study to develop and validate risk assessment models for diabetes mellitus and pre-diabetes mellitus. The study is reported following the guidelines of the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) statement²³.

Outcome and risk factors

The outcome is newly detected diabetes mellitus and pre-diabetes mellitus by blood tests. According to the WHO²⁴, ADA²⁵ and the Hong Kong Reference Framework for Diabetes Care for Adults in Primary Care Setting²⁶, pre-diabetes mellitus was defined as a fasting plasma glucose of 6.1–6.9 mmol/L or HbA1c of 5.7–6.4%, and diabetes mellitus was defined as a fasting plasma glucose higher or equal to 7.0 mmol/L or HbA1c higher or equal to 6.5%.

We included all available socio-demographics, lifestyle factors, and non-laboratory clinical parameters in the model development. The socio-demographics included age and sex. Lifestyle factors included smoking, alcohol consumption, physical activity, sleep duration and quality, and dietary habits. Alcohol consumption was measured using the Alcohol Use Disorders Identification Test Alcohol Consumption Questions (AUDIT-C) which is a 3-item screening tool based on the WHO AUDIT²⁷. Physical activity was measured using the WHO Global Physical Activity Questionnaire²⁸. Sleep was assessed using six items, including sleep duration, self-assessed insufficient sleep, self-assessed overall sleep quality, and the presence of sleep disturbance (i.e., difficulty in falling asleep, intermittent awakening, early morning awakening). Dietary habit was assessed using daily fruit and vegetable consumption (standard servings per day), and monthly eat-out frequency. Clinical parameters included systolic blood pressure (SBP), diastolic blood pressure (DBP), body mass index (BMI), waist circumference, and waist-to-hip ratio (WHR). The detailed definition and measurement of risk factors can be found in the PHS

2014/15 Report²². No missing value was present in the study dataset.

Statistical analysis

Descriptive statistics on the characteristics of the subjects were tabulated by groups of diabetes mellitus, pre-diabetes mellitus, and normal glycemia. The differences of each risk factor among the different glycemia groups (DM/pre-DM/normal glycemia) were compared using ANOVA for continuous variables and Chi-square for categorical variables. Post hoc pairwise comparison *P* values were adjusted by the Bonferroni method.

The study sample was randomly split with a ratio of two-to-one for the development ($n = 1238$) and validation ($n = 619$) of the risk models. To cross-validate the risk factors and to optimize the performance, we used both traditional logistic regression (LR) and machine learning (ML) algorithms to develop the risk models from the data of the development sample. Multicollinearity of the predictors were diagnosed using variance inflation factors (VIF) based on the full logistic regression model. A VIF > 5 indicates the existence of multicollinearity and greater than 10 indicates severe multicollinearity²⁹. The LR model was developed using the Akaike information criterion (AIC) based bidirectional stepwise multivariable logistic regression. The combination of risk factors that achieved the lowest AIC value was included in the model. Quadratic terms of the included risk factors, as well as their interactions with age, were also evaluated based on their statistical significance to improve the fitting of the LR model. The final LR model was established by combining the coefficients of the risk factors and the logistic function. The ML model was developed using Extreme Gradient Boosting (Xgboost)³⁰. The hyper-parameters of Xgboost were determined by a 5-fold cross-validation grid search. The predicted probability of the Xgboost model was calibrated using the isotonic method to improve the results³¹. The Shapley Additive Explanations (SHAP)³² method was used to evaluate the importance of the risk factors and to show the nonlinear relationship and interactive effects inside the ML model, by way of calculating the marginal contributions of the risk factors. Besides, the Boruta algorithm³³ was used to select statistically the most important risk factors without pre-defining an importance threshold, by introducing randomized variables (also referred to as shadow variables). The rationality of the ML model was reviewed by clinical experts (CLK, EYTY, EYTY), considering the clinical significance of the nonlinear effect of the risk factors. The optimal risk cut-offs for the LR and ML models were determined by Youden's index³⁴.

The performance of the LR and ML models was tested on the validation sample. The discrimination power was evaluated using the area under the curve of the receiver-operating characteristic curve (AUC-ROC) and the precision-recall curve (AUC-PR). The AUC-ROC ranges from 0.5 to 1, where 0.7 to 0.8 is considered good and more than 0.8 is considered excellent³⁵. The AUC-PR is a performance metric measuring the model's ability to detect positive cases, which is a recommended

evaluation when the proportion of positive cases is small³⁶. A higher AUC-PR indicates better performance but there is no agreed standard. The confidence intervals of AUC-ROC and AUC-PR were estimated using bootstrap. The sensitivity (recall), specificity, positive predictive value (PPV or precision), the negative predictive values (NPV) at different risk thresholds were calculated. Model calibration was assessed by calibration plots^{37,38} and the Hosmer-Lemeshow test, to measure how well the predicted risk agreed with the observed event rate. Two existing diabetes mellitus risk models specific for the Chinese population, the NCDRS¹⁸ and the NDS¹⁹, and the screening recommendation by the Hong Kong Reference Framework for Diabetes Care for Adults in Primary Care Settings¹² were also applied to the validation sample for performance comparison of detecting diabetes mellitus and pre-diabetes mellitus, and diabetes mellitus only. The AUC-ROCs of different models were compared using DeLong's test³⁹, and the AUC-PRs were compared using a bootstrap-based test⁴⁰ with MedCalc 19.8. Net reclassification improvement (NRI) and integrated discrimination improvement (IDI) were also used to compare different models, on the aspects of changes in risk classification and changes in risk difference between events and non-events, respectively⁴¹. A NRI and IDI significantly greater than zero indicate a better performance of the updated model.

All significance tests were two-tailed, with a significance level at a *P*-value of <0.05. Data analyses were conducted using R 3.5.1 and Python 3.6.

RESULTS

Among the 1,857 subjects, 47.7% were male and the mean \pm standard deviation age was 40.7 ± 15.5 years old. Subject characteristics by glycemic groups are shown in Table 1. The prevalence of new diabetes mellitus and pre-diabetes mellitus as detected by blood tests were 3.77% ($n = 70$) and 11.31% ($n = 210$), respectively. The total prevalence of newly detected diabetes mellitus and pre-diabetes mellitus was 15.08% ($n = 280$).

Development of DM and pre-DM risk assessment models

The multicollinearity diagnosis results showed that the highest VIF of the possible predictors came from waist circumference at 4.41, indicating no severe multicollinearity existed. The results from the LR risk model are presented in Table 2, showing seven significant risk factors, including age, BMI, WHR, smoking status, sleep duration, vigorous recreational activity time per week, and fruit consumption per week. Age showed a significant non-linear effect on the outcome (odds ratio of age²: 0.999 [0.998, 1.000]), in that the risk of new diabetes mellitus and pre-diabetes mellitus reached a peak at the age of 74 years old. An age-dependent effect of sleep duration on the risk of new diabetes mellitus and pre-diabetes mellitus was observed (odds ratio of interaction term: 1.015 [1.004, 1.027]). Specifically, the effect of short sleep duration decreased with age. The final function of the LR model is: $1/(1 + e^{-(0.0854 * \text{Age} +$

Table 1 | Subject characteristics overall and by glycemic status (n = 1,857)

Characteristic	Overall (n = 1,857)	DM (n = 70)	Pre-DM (n = 210)	Normal glycemia (n = 1,577)
Demographics				
Age, years	40.70 ± 15.48	55.46 ± 12.29 ^b	53.22 ± 12.87 ^b	38.37 ± 14.76
Sex, male	885 (47.66%)	46 (65.71%) ^{a,b}	96 (45.71%)	743 (47.11%)
Smoking status (current smoker)	226 (12.17%)	17 (24.29%) ^b	37 (17.62%) ^b	172 (10.91%)
Clinical parameters				
SBP, mmHg	115.77 ± 17.36	127.91 ± 18.65 ^b	124.46 ± 18.33 ^b	114.08 ± 16.61
DBP, mmHg	76.60 ± 10.39	80.67 ± 11.73 ^b	79.39 ± 9.72 ^b	76.04 ± 10.32
BMI, kg/m ²	23.03 ± 3.77	26.18 ± 4.78 ^{a,b}	24.87 ± 3.71 ^b	22.64 ± 3.59
WHR	0.84 ± 0.07	0.91 ± 0.06 ^{a,b}	0.88 ± 0.07 ^b	0.83 ± 0.07
Waist circumference, cm	79.68 ± 10.65	89.14 ± 10.26 ^{a,b}	84.72 ± 9.70 ^b	78.58 ± 10.38
Drinking habit				
Drinking frequency				
Never	528 (28.43%)	23 (32.86%)	70 (33.33%)	435 (27.58%)
Monthly or less	1003 (54.01%)	37 (52.86%)	115 (54.76%)	851 (53.96%)
Twice a month or more	326 (17.56%)	10 (14.29%)	25 (11.90%)	291 (18.46%)
Alcohol consumption each time, unit	2.00 ± 2.90	1.92 ± 2.60	1.53 ± 1.98 ^b	2.06 ± 3.01
Harmful drinking frequency				
Never	1712 (92.19%)	63 (90.00%)	200 (95.24%)	1449 (91.88%)
Less than monthly	93 (5.01%)	3 (4.29%)	7 (3.33%)	83 (5.26%)
Monthly or more	52 (2.8%)	4 (5.71%)	3 (1.43%)	45 (2.86%)
AUDIT score	2.07 ± 2.64	2.31 ± 3.08	1.75 ± 2.26 ^b	2.10 ± 2.67
Sleeping				
Sleeping duration, hour/day	6.90 ± 1.19	6.76 ± 1.18	6.69 ± 1.34 ^b	6.93 ± 1.16
Days of poor sleep in last month	7.08 ± 9.16	6.93 ± 8.72 ^b	7.70 ± 9.96	7.01 ± 9.07
Self-conceived sleep quality				
Good	1044 (56.22%)	31 (44.29%)	113 (53.81%)	900 (57.07%)
Fair	618 (33.28%)	33 (47.14%)	76 (36.19%)	509 (32.28%)
Poor	195 (10.50%)	6 (8.57%)	21 (10.00%)	168 (10.65%)
Difficulty in falling asleep yes	609 (32.79%)	26 (37.14%)	77 (36.67%)	506 (32.09%)
Intermittent awakenings, yes	640 (34.46%)	36 (51.43%) ^b	87 (41.43%) ^b	517 (32.78%)
Early morning awakening yes	537 (28.92%)	32 (45.71%) ^b	77 (36.67%) ^b	428 (27.14%)
Physical activity				
Vigorous recreational activity time, min/week	37.18 ± 111.41	16.00 ± 57.77	14.36 ± 57.53 ^b	41.15 ± 118.01
Moderate recreational activity time, min/week	55.47 ± 125.56	67.86 ± 115.99	77.24 ± 148.98 ^b	52.02 ± 122.28
Vigorous work time min/week	61.69 ± 349.66	68.57 ± 330.42	106.29 ± 454.51 ^b	54.57 ± 331.18
Moderate work time min/week	121.43 ± 424.13	158.36 ± 497.18	133.57 ± 485.05	117.30 ± 409.31
Travel to and from places min/week	456.29 ± 479.37	443.43 ± 386.62	443.79 ± 457.86	457.88 ± 484.76
Sedentary behavior time min/week	2905 ± 1127	2871 ± 1113	2838 ± 1188	2916 ± 1119
Overall energy expenditure MET/week	3313 ± 4154	3355 ± 3934	3580 ± 5093	3275 ± 4024
WHO PA level (physically active)	1640(88.31%)	64 (91.43%)	181 (86.19%)	1395 (88.46%)
Diet				
Fruit consumption servings/week	32.33 ± 35.53	34.29 ± 42.98	31.09 ± 26.74	32.41 ± 36.20
Vegetable consumption servings/week	60.62 ± 60.34	67.87 ± 80.81	59.48 ± 48.34	60.45 ± 60.75
Eat-out frequency times/month	31.13 ± 20.91	24.79 ± 22.02 ^b	26.07 ± 21.06 ^b	32.09 ± 20.71

All characteristics are expressed in either number (percentage) or mean (SD). Post-hoc pairwise comparisons among groups of DM/pre-DM/normal glycemia were conducted using *t*-test or Chi-square test with *P* values adjusted by Bonferroni method. ^aThe difference between diabetes mellitus and pre-diabetes mellitus groups was statistically significant (*P* < 0.05). ^bThe difference between diabetes mellitus group or pre-diabetes mellitus group and normal glycemia group was statistically significant (*P* < 0.05). AUDIT score, alcohol use disorder identification test score; BMI, body mass index; DBP, diastolic blood pressure; DM, diabetes mellitus; MET, metabolic equivalent; PA, physical activity; pre-DM, pre-diabetes; SBP, systolic blood pressure; WHR, waist to hip ratio.

Table 2 | Diabetes mellitus and pre-diabetes mellitus risk factors of prediction model developed by logistic regression ($N = 1,238$)

	Coefficient	OR (95%CI)	P value
Age, years	0.0854	1.0891 (0.9768, 1.2143)	0.124
BMI, kg/m ²	0.1251	1.1332 (1.0739, 1.1959)	<0.001
WHR	0.2295	1.2579 (0.9110, 1.7370)	0.163
Smoker (ref. non-smoker)	0.5562	1.7440 (1.0882, 2.7952)	0.021
Sleeping duration, hour/day	-0.9718	0.3784 (0.1989, 0.7200)	0.003
Vigorous recreational activity time, min/week	-0.0026	0.9974 (0.9948, 1.0000)	0.047
Fruit consumption, servings/week	-0.0041	0.9959 (0.9905, 1.0013)	0.136
Age ²	-0.0012	0.9988 (0.9979, 0.9997)	0.009
Age*Sleep duration	0.0152	1.0153 (1.0037, 1.0270)	0.009
Constant	-6.0591		

The risk model was developed using AIC-based stepwise multivariable logistic regression. Variables that could significantly improve the model's goodness of fit measure were selected. The unit of change of WHR is 0.1, and the unit of changes of all other parameters is 1. BMI, body mass index; CI, confidence interval; DM, diabetes mellitus; pre-DM, pre-diabetes; WHR, waist to hip ratio.

$0.1251 \times \text{BMI} + 2.2947 \times \text{WHR} + 0.5562 \times \text{Smoker} - 0.9718 \times \text{Sleep duration} - 0.0026 \times \text{Vigorous recreational activity time} - 0.0041 \times \text{Fruit consumption} - 0.0012 \times \text{Age}^2 + 0.0152 \times \text{Age} \times \text{Sleep duration} - 6.0591$).

Using data of the same subjects ($N = 1,238$), the importance ranking of the risk factors and variable selection result of the ML model developed by Xgboost are presented in Figure 1. Eight risk factors, including age, BMI, WHR, SBP, waist circumference, sleep duration, smoking status, and vigorous recreational activity time per week, were selected by the Boruta method for inclusion in the final ML model. The relationships between each risk factor and the risk of new diabetes mellitus and pre-diabetes mellitus are shown in Figure 2. The effect of important interactions between risk factors is shown in Figure S1 with color scale rulers. The effect of age increased sharply from the age of 35 years and peaked at the age of 60 years. The BMI showed a significant interaction with age, in that after the age of 50, the effect of age on diabetes mellitus and pre-diabetes mellitus among people with a higher BMI were stronger than those with a low BMI.

Sleep duration showed a non-linear relationship with the risk of new diabetes mellitus and pre-diabetes mellitus, where individuals with sleep duration of 7 to 8 hours showed the lowest risk. Vigorous recreational activity time per week showed a protective effect especially in the elderly, and the relationship was most prominent from 0 to 120 min per week.

Validation of DM and pre-DM risk assessment models

The ROC and PR curves evaluating the discrimination of the LR, ML, and two existing models on diabetes mellitus and pre-diabetes mellitus cases and diabetes mellitus cases only are shown in Figure 3. For the detection of diabetes mellitus and pre-diabetes mellitus, the ML model showed the best discrimination with an AUC-ROC of 0.822 [0.779, 0.863] and AUC-PR of 0.496 [0.391, 0.602], which was significantly higher (P -value < 0.05) than those of the LR model (AUC-ROC = 0.812

[0.769, 0.853], AUC-PR = 0.448 [0.361, 0.535]), NCDRS (AUC-ROC = 0.784 [0.739, 0.828], AUC-PR = 0.364 [0.276, 0.451]), and NDS (AUC-ROC = 0.786 [0.740, 0.831], AUC-PR = 0.378 [0.270, 0.487]). The AUC-ROC and AUC-PR of the LR model were significantly higher than those of NCDRS and NDS ($P < 0.05$). The NRI and IDI of the ML model over the LR model were both significantly greater than zero (NRI = 0.27 [0.13, 0.42], IRI = 0.07 [0.04, 0.11]), indicating a better performance of the ML model than the LR model. For the detection of diabetes mellitus only, the ML model had the highest AUC-ROC of 0.837 [0.784, 0.888] and AUC-PR of 0.178 [0.058, 0.298], and both the ML and LR models had a significantly better discrimination power than the NCDRS and NDS. To avoid that the results were due to chance, data splitting (random splitting of the development and validation sample at 2:1) was repeated 20 times and the performance of the risk models remained largely unchanged (Table S1).

The optimal risk threshold to detect diabetes mellitus and pre-diabetes mellitus identified by Youden's index was 12.7% for the ML model and 11.0% for the LR model. The sensitivity (recall), specificity, PPV (precision), and NPV of the ML model and the LR model at different risk thresholds are listed in Table 3. Using the same risk threshold, the LR model showed better sensitivity and NPV, whereas the ML model showed a higher specificity and PPV. Using the PHS 2014/15 the prevalence of pre-diabetes mellitus and diabetes mellitus at 15%, the ML model had a sensitivity of 72.4%, a specificity 77.9%, PPV 38.2%, and NPV 93.8%; and the LR model had a sensitivity of 77.6%, a specificity 68.1%, PPV 31.4%, and NPV 94.2%. The corresponding specificity, PPV, and NPV of the risk models by sensitivity levels are also listed in Table 3.

The calibration plots of the LR and ML risk models are shown in Figure 4. Both the ML and LR models showed good calibration, as the difference between the predicted risk and the observed risk was not statistically significant (H-L test P -value > 0.05). The LR model tended to underestimate the risk

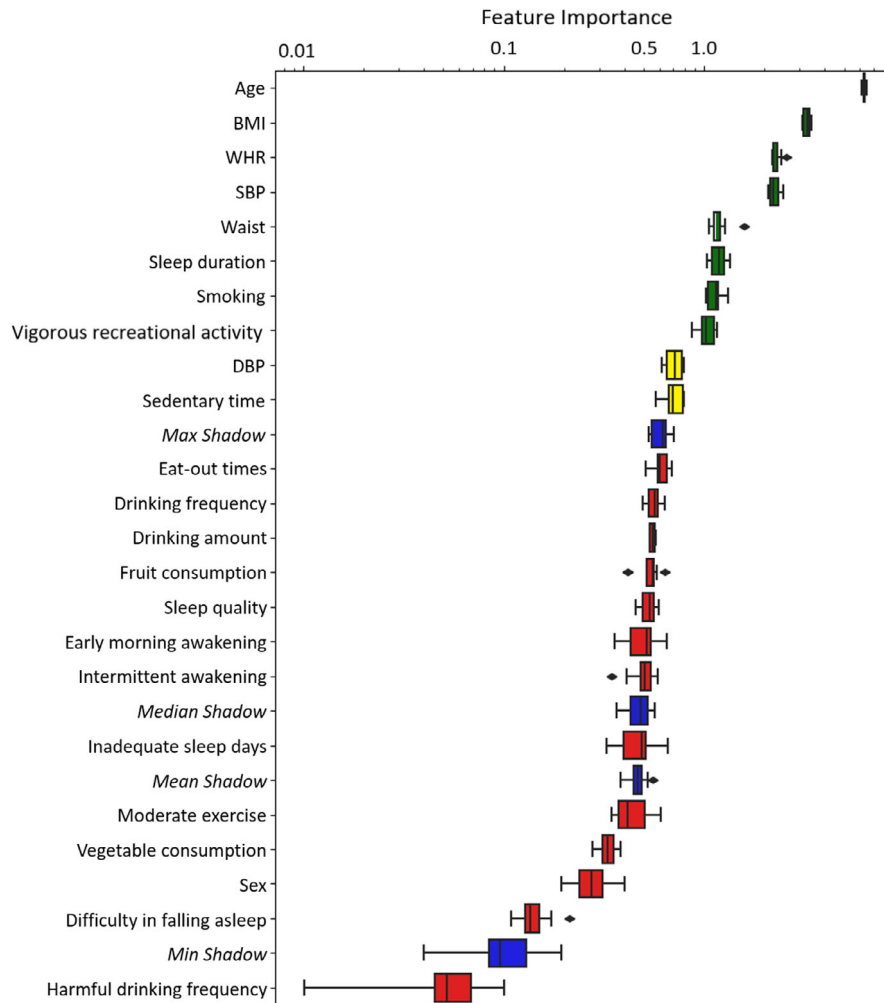


Figure 1 | Diabetes mellitus and pre-diabetes mellitus risk factor (feature) selection and importance ranking by ML modeling ($N = 1,238$). BMI, body mass index; DBP, diastolic blood pressure; DM, diabetes mellitus; ML, machine learning; pre-DM, pre-diabetes; SBP, systolic blood pressure; WHR, waist to hip ratio. Feature selection was conducted using Boruta algorithm, based on the feature importance calculated by SHAP. Blue bars indicate the randomized variables (shadow variables). Variables with significantly higher importance than the randomized variables are considered to be important. Green bars indicate the important risk factors. Yellow bars indicate the marginally important risk factors. Red bars indicate the unimportant risk factors.

when the risk was <0.2 (20%), and the ML model tended to underestimate the risk when the risk was more than 0.2 (20%). At the bottom of each calibration plot, a histogram of the number of subjects at different predicted risks shows that most subjects had a risk between 0 and 0.2. Hence overall the ML model is more resistant to misclassification.

Deployment of the risk models

The risk models developed in this study have been deployed as a computerized calculator as displayed in Figure S2. The calculator can estimate the absolute risk (0–100%) of diabetes mellitus and pre-diabetes mellitus from the input information on the risk factors, using the LR model and ML model, respectively. The clinician can decide on the need of further blood

tests based on the estimated risk and the associated sensitivity, specificity, positive and negative predictive values (Table 3). The risk assessment calculator is available online with detailed installation and operation instructions (<https://github.com/dongdongdongdwn/Non-laboratory-DM-and-pre-DM-risk-model-for-case-detection-in-Chinese-population.git>).

DISCUSSION

The current study has demonstrated the utility of diabetes mellitus and pre-diabetes mellitus risk assessment models that include only non-laboratory-based risk factors that are available in routine clinical practice. It has the strength of using data from a sample representative of the general population, which is generalizable and most applicable to primary care. Another

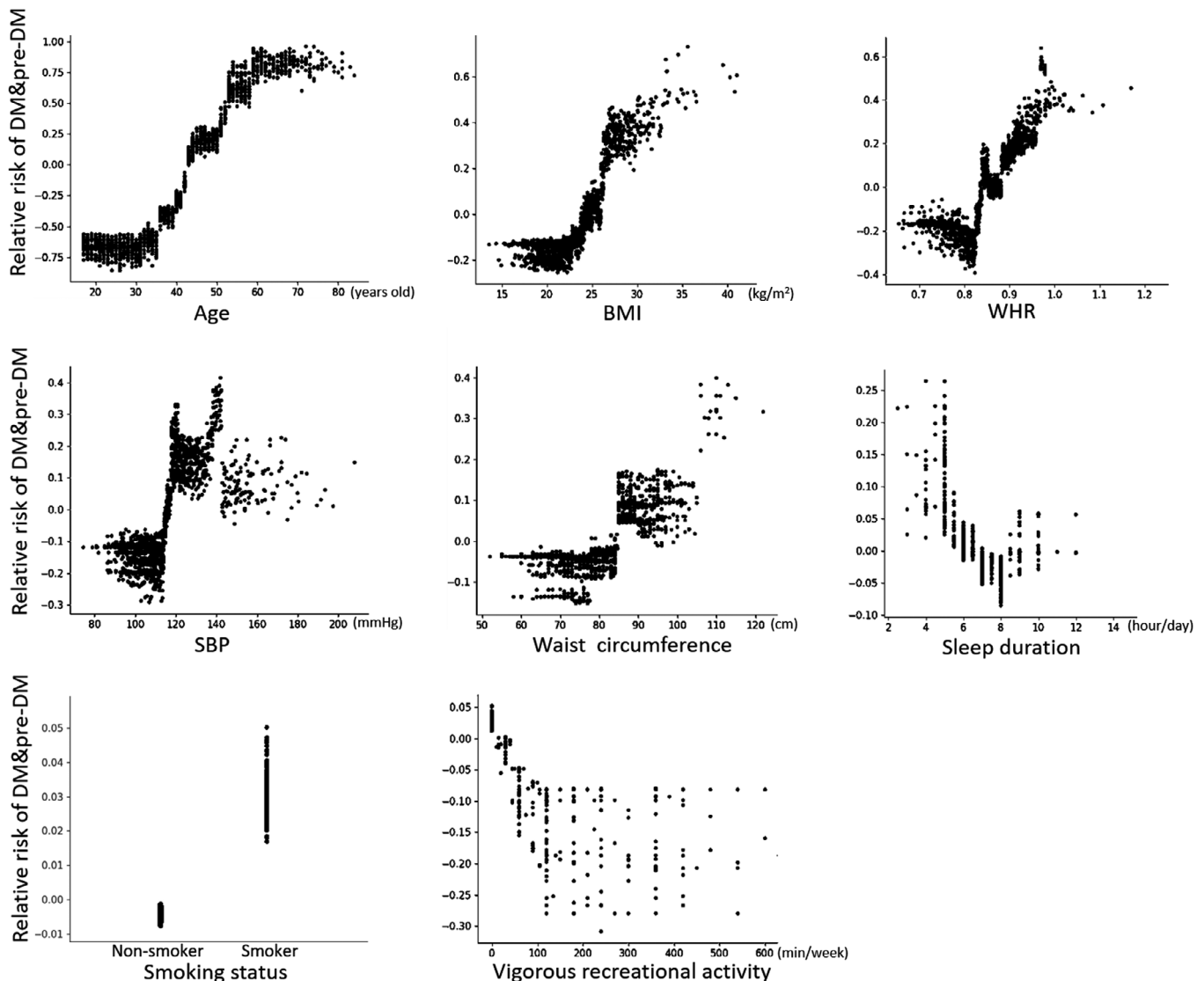


Figure 2 | Relationship between risk factors (feature) and relative risk of new diabetes mellitus and pre-diabetes mellitus by ML modeling ($N = 1,238$). BMI, body mass index; DBP, diastolic blood pressure; DM, diabetes mellitus; ML, machine learning; pre-DM, pre-diabetes; SBP, systolic blood pressure; WHR, waist to hip ratio. The SHAP method was used to interpret the fitting result of the ML model. Nonlinear relationships between each risk factor (x-axis) and the relative risk of DM and pre-DM to the study population level (y-axis) are shown.

strength is the use of both LR and ML methods to develop the risk assessment models showed largely similar risk factors, supporting the validity of the results. The LR risk model and the ML model both showed better discrimination power than the two existing diabetes mellitus risk scoring models for the detection of diabetes mellitus and pre-diabetes mellitus and diabetes mellitus only. They were also more accurate than the screening criteria recommended by the Hong Kong Reference Framework for Diabetes Care for Adults in Primary Care Settings. Considering the calibration and the prevalence of an undiagnosed diabetes mellitus and pre-diabetes mellitus found in the PHS of

15%, the ML model is likely to be more resistant to misclassification bias.

In addition to the well-known risk factors of diabetes mellitus (age, BMI, WHR, SBP, waist circumference, and smoking status), this study also found that sleep duration and vigorous recreational activity time per week were significant risk factors of new diabetes mellitus and pre-diabetes mellitus, both of which are important predictors identified in the LR and ML models. Kengne *et al.*⁴² summarized existing non-invasive risk models of type 2 diabetes mellitus and had similarly found that age, smoking, family history, BMI, waist circumference,

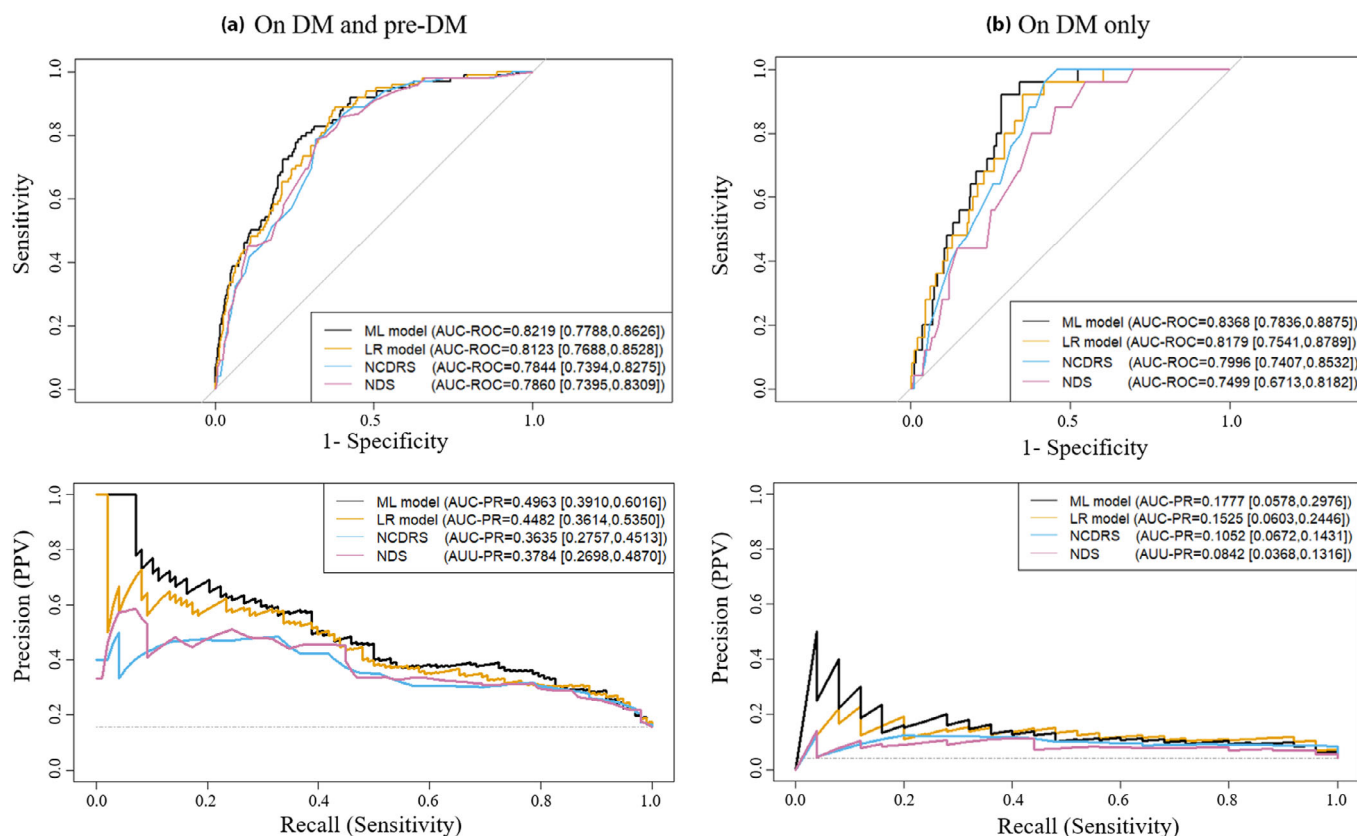


Figure 3 | ROC and PR curves of risk prediction models to detect new diabetes mellitus and pre-diabetes mellitus (DM only) on the validation sample ($N = 619$). AUC, area under curve; LR, logistic regression; ML, machine learning; NCDRS, the New Chinese Diabetes Risk Score; NDS, non-invasive diabetes score; PPV, positive predictive value; PR, precision-recall; ROC, receiver-operating characteristic. 95% CIs were calculated using bootstrap. (a) For diabetes mellitus and pre-diabetes mellitus detection, the ML model showed significantly better AUC-ROC (DeLong's test P value < 0.05) and AUC-PR (bootstrap-based test P value < 0.05) than those of LR model, NCDRS and NDS. The LR model showed significantly better AUC-ROC and AUC-PR than NCDRS and NDS. Continuous net reclassification improvement (NRI) and integrated discrimination improvement (IDI) of the ML model beyond the LR model were 0.27 [0.13, 0.42] and 0.07 [0.04, 0.11], respectively, both significantly higher than 0 ($P < 0.05$). (b) The ML model showed significantly better AUC-ROC (DeLong's test P value < 0.05) than the LR model, NCDRS and NDS. The LR model showed significantly better AUC-ROC than NCDRS and NDS (DeLong's test P value < 0.05). The ML model and LR model both showed significantly higher AUC-PR than NCDRS and NDS, but the difference of AUC-PR between the ML model and the LR model was not significant.

hypertension, and physical activity were the most commonly used risk factors. Our risk models also considered the nonlinear effect of these well-known risk factors by using transformation and interaction terms in LR, and by ML to improve model performance over existing models.

It is interesting to note that SBP was not a significant predictor in the LR model but was one of the most important risk factors in the ML model. The interrelationships among risk factors are complex and linear adjustment of other covariates might dilute the actual nonlinear effect of some factors in LR model. The ML model analysis showed that the risk of diabetes mellitus and pre-diabetes mellitus increased sharply above a SBP of 120 mmHg, suggesting the threshold of SBP is 120 mmHg for the risk of diabetes mellitus and pre-diabetes mellitus. People without hypertension but with elevated SBP of

more than 120 mmHg should be targeted for diabetes mellitus and pre-diabetes mellitus screening.

The BMI, WHR, and waist circumference are all frequently used indicators of obesity. The ML model included all three and the LR model included two (BMI and WHR) of them, which raised the issue of multicollinearity. The VIF of these three parameters in regression were all < 5 , indicating no significant multicollinearity existed. We further carried out pairwise correlation analysis of these three parameters (Figure S3), and found BMI, WHR, and waist circumference were linearly correlated, but they were not redundant (Pearson relationship < 0.7). The stepwise LR model selected BMI and WHR but not waist circumference, indicating that WHR may be a stronger predictor of diabetes mellitus and pre-diabetes mellitus than waist circumference when only the linear effect was considered. On the

Table 3 | Sensitivity, specificity, PPV, and NPV of diabetes mellitus and pre-diabetes mellitus risk models at different risk thresholds and at different sensitivity levels ($N = 619$)

Risk threshold	Model	Sensitivity	Specificity	PPV	NPV
Models' performance at different risk thresholds (10%/15%/20%/25%)					
Best threshold	LR model	0.888	0.622	0.306	0.967
LR 11.0%, ML 12.7%	ML model	0.786	0.739	0.362	0.948
10%	LR model	0.888	0.601	0.295	0.966
	ML model	0.827	0.653	0.309	0.952
15%	LR model	0.776	0.681	0.314	0.942
	ML model	0.724	0.779	0.382	0.938
20%	LR model	0.663	0.764	0.346	0.923
	ML model	0.571	0.821	0.376	0.911
25%	LR model	0.571	0.816	0.368	0.910
	ML model	0.500	0.868	0.415	0.902
Models' performance at different sensitivity levels (0.9/0.8/0.7)					
7.9%	LR model	0.900	0.557	0.278	0.970
8.1%	ML model	0.900	0.574	0.286	0.971
16/50	NCDRS	0.900	0.493	0.254	0.970
10/50	NDS	0.900	0.501	0.255	0.967
13.5%	LR model	0.800	0.664	0.311	0.948
11.3%	ML model	0.800	0.673	0.345	0.951
21/50	NCDRS	0.800	0.633	0.298	0.951
19/50	NDS	0.800	0.649	0.299	0.944
18.7%	LR model	0.700	0.747	0.343	0.931
16.2%	ML model	0.700	0.787	0.383	0.934
23/50	NCDRS	0.700	0.699	0.302	0.924
22/50	NDS	0.700	0.708	0.309	0.925
HK reference framework for diabetes care for adults in primary care setting					
		0.942	0.353	0.227	0.968

Optimal risk cutoffs of LR model and ML model were determined by Youden's index. As NCDRS and NDS only provides risk scores instead of corresponding absolute risk in percentage, the indexes by risk thresholds cannot be calculated for these two models. The HK Reference Framework for Diabetes Care for Adults in Primary Care Setting is a risk factor-based screening criteria and neither risk estimation nor risk score is provided, hence only a set of sensitivity, specificity, PPV, and NPV is presented in the table. LR, logistic regression; ML, machine learning; NCDRS, the New Chinese diabetes risk score; NDS, non-invasive diabetes score; NPV, negative predictive value; PPV, positive predictive value.

other hand, the ML model identified these three obesity indicators were all significant risk factors, and the inclusion of all three indicators provides a more accurate risk assessment. The independent nonlinear effects of these three parameters after adjustment (Figure 2) were in line with clinical experience and published literature^{43,44}. This implies that the ML model can extract additional predictive information from some predictors that the linear model cannot detect. There is no consensus on which of these three parameters is the best indicator of obesity⁴⁵. Some studies have verified that waist circumference and

WHR can provide extra information on diabetes mellitus incidence in addition to BMI^{46,47}. These parameters may provide predictive power singly or in combination for different individual patients. In addition, it seems remarkable that the ML model showed a dramatic increase in diabetes mellitus and pre-diabetes mellitus risk at a waist circumference of 85 cm, which is consistent with the waist circumference thresholds observed in other Asian populations^{43,48,49}. For example, a Japanese cohort identified that a waist circumference of 85/80 cm for male/female was the best cut-off for metabolic syndrome⁴³. A waist circumference of around 85 cm, which is much lower than the recommended 102 cm for Western populations⁵⁰, should be a more appropriate cut-off point for Asians to stratify the risk of diabetes and other metabolic disorders.

In addition to conventional risk factors, sleep duration and vigorous recreational activity time per week were identified as significant risk factors of diabetes mellitus and pre-diabetes mellitus in both the LR and ML models. Both predictors are modifiable lifestyle factors, hence their importance in diabetes mellitus risk intervention. With the ML model, sleep duration (as a continuous variable) showed a U-shaped relationship where subjects with 7–8 h of sleep per day showed the lowest risk of diabetes mellitus and pre-diabetes mellitus. We further tested the statistical association between sleep duration levels (<7 h, 7–8 h, >8 h) and risk of diabetes mellitus and pre-diabetes mellitus in our subjects and found, as shown in Table S2, the effect of excessive sleep (>8 h) did not reach statistical significance ($P > 0.05$), which could be related to the large variance in a small sub-sample. A meta-analysis conducted from 11 prospective studies found that when compared with the sleep duration category of 7–8 h per day, both insufficient and excessive sleep duration were associated with an increased risk of type 2 diabetes mellitus⁵¹. Given all these, sleep duration should be considered in diabetes mellitus and pre-diabetes mellitus risk assessment. Physical activity (PA) is a well-recognized risk factor of diabetes mellitus and has been included in the ADA risk model¹³, AUSDRISK¹⁵ and CAN-RISK¹⁶ models, where physical activity is measured by self-reported time on total physical activities. Our study measured physical activity using the WHO's Global Physical Activity Questionnaire (GPAQ)²⁸, which enquires on a detailed account of all activities at work, during travel to and from places, and on recreation. It should be noted that only vigorous recreational activity time per week was a significant risk factor of diabetes mellitus and pre-diabetes mellitus, whereas other types of physical activity, overall energy expenditure in METs, and physical activity levels according to WHO recommendations were insignificant. A Japanese cohort also found that only vigorous-intensity leisure-time exercise was associated with risk of type 2 diabetes mellitus, whereas the associations were insignificant for moderate-intensity exercise and occupational physical activity⁵². These results were further confirmed by a Chinese cohort study⁵³ and a multi-ethnic cohort study⁵⁴. In addition, the nonlinear trend identified by our ML model between vigorous

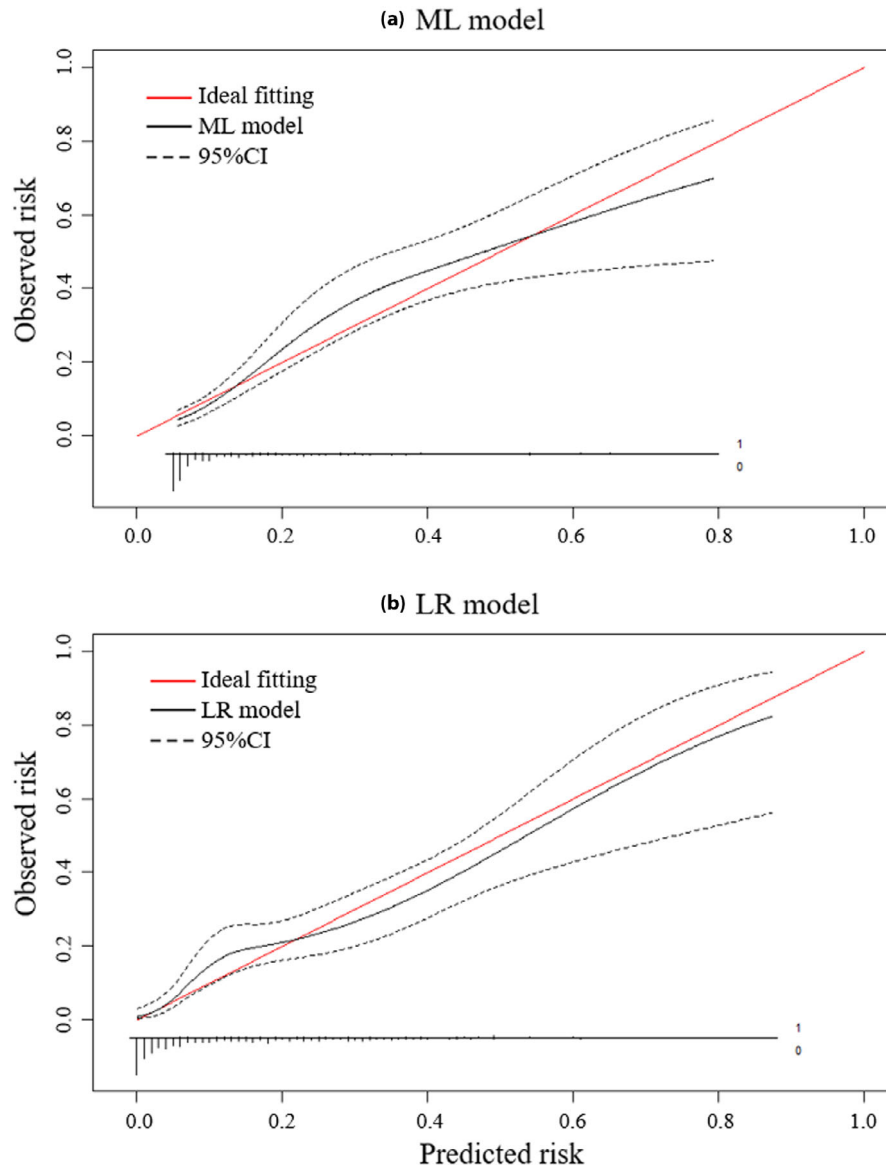


Figure 4 | Calibration plots of risk prediction models to detect new diabetes mellitus and pre-diabetes mellitus on the validation sample ($N = 619$). Hosmer-Lemeshow test results showed the difference between predicted risk and observed risk was not significant ($P > 0.05$) for both the LR and ML models. The x-axis is the predicted risk of diabetes mellitus and pre-diabetes mellitus, and the y-axis is the observed risk of diabetes mellitus and pre-diabetes mellitus. The curves were fitted based on restricted cubic splines. At the bottom of the graphs, histograms of the predicted risks are shown for the subjects with (1) and without (0) diabetes mellitus and pre-diabetes mellitus. Since NCDRS and NDS only provides risk scores instead of corresponding absolute risk in percentage, their calibration cannot be evaluated. Since the models were to estimate the risk of diabetes mellitus and pre-diabetes mellitus, hence the calibration on DM only were not carried out.

recreational activity time and risk of diabetes mellitus and pre-diabetes mellitus was in line with the finding from a meta-analysis of ten cohort studies, in that more pronounced dose-response reduction in the risk of diabetes mellitus was observed at vigorous recreational activities of 0–2 h per week⁵⁵. Taken together, focusing on the assessment of vigorous recreational activity time per week might be more sensitive for diabetes

mellitus and pre-diabetes mellitus risk assessment, which would also enhance the acceptability and efficiency in data collection in primary care.

Given that the estimated risk of new diabetes mellitus and pre-diabetes mellitus is a continuous value ranging between 0 and 100%, the risk threshold to be adopted for case detection has to consider the trade-off between sensitivity and precision

under different circumstances. At the same sensitivity level, the ML model showed better specificity and precision than the LR model. For example, if 80% of cases of diabetes mellitus and pre-diabetes mellitus need to be detected successfully (sensitivity = 80%), the precision of the ML model is 0.345, corresponding to a number-needed-to-screen of 2.9 to identify one case of diabetes mellitus and pre-diabetes mellitus, whereas the number for the LR model (precision = 0.311) is 3.2, for NCDRS (precision = 0.298) and NDS (precision = 0.299) is 3.4. This difference can be significant when screening has to be applied to a large population on an ongoing basis in primary care.

The vast majority of existing diabetes mellitus risk assessment models were developed using regression-based methods⁵⁶, which are limited in their ability to handle complex relationships and may lead to suboptimal results. The ML model developed in this study showed outstanding discrimination and calibration, and surpassed the developed LR model and the two existing models (NCDRS and NDS). ML algorithms can provide more accurate risk assessments due to their powerful fitting ability but they have been criticized for lack of transparency⁵⁷. In this study, we showed the SHAP method could improve the interpretability of the ML model by quantifying and visualizing the nonlinear and interactive effects of each risk factor. Incorporating the review of clinicians based on their experience and knowledge, the reliability and usability of the ML models were substantially improved, ensuring the models developed in this study have the potential to be integrated into type 2 diabetes mellitus screening and prevention in routine clinical practice.

This study has several limitations. First, some well-known risk factors, such as a family history of diabetes mellitus and a history of gestational diabetes mellitus could not be included because they were not collected in the PHS 2014/15. Second, the validation was carried out on a sample from the same population, therefore, further validation on an external sample in primary care should be carried to establish its validity in clinical practice. Third, due to the exclusion criteria, the model may not be generalizable to individuals with a known diagnosis of hypertension, CVD, cancer, renal disease, or anemia.

CONCLUSION

Using a representative sample of the Chinese general population, this study developed a non-laboratory-based risk assessment models to detect undiagnosed diabetes mellitus and pre-diabetes mellitus in Chinese adults using both a classical statistical method and an interpretable machine learning method. Besides conventional diabetes mellitus risk factors, sleep duration of less than 7 h and vigorous recreational activity time of less than 120 min per week were found to be significant modifiable risk factors of diabetes mellitus and pre-diabetes mellitus, which should be included in future risk assessment models as well as interventions to prevent diabetes mellitus and pre-

diabetes mellitus. The new models developed in this study had excellent performance with ROC-AUC >0.8 in the validation sample, which was better than existing risk models and the Chinese-specific Reference Framework for the detection of diabetes mellitus and pre-diabetes mellitus. Subject to confirmation by external validation in primary care, the models can be incorporated in the electronic medical record system or made available as a mobile application to facilitate opportunistic case detection of diabetes mellitus and pre-diabetes mellitus in primary care. Another potential application is for patient activation to self-monitor their own risk of diabetes mellitus and pre-diabetes mellitus.

ACKNOWLEDGMENTS

The authors wish to acknowledge the Department of Health of the Government of the Hong Kong Special Administrative Region for the provision of study data. This work was supported by Health and Medical Research Fund (HMRF ref No. 17181641), the Food and Health Bureau, The Government of the Hong Kong Special Administrative Region. The funding organization did not play any role in this study.

DISCLOSURE

The authors declare no conflict of interest.

Approval of the research protocol: Ethics approval was granted by the Institutional Review Board (IRB) of the University of Hong Kong/Hospital Authority Hong Kong West Cluster on 24 December 2019 (Reference no. UW 19-831).

Informed consent: Hong Kong Department of Health obtained informed consent from all individual participants included in the study, and approved the usage of the data for this study.

Registry and the registration no. of the study/trial: US Clinical-Trial.gov: NCT04881383, May 11, 2021; HKU clinical trials registry: HKUCTR-2808, December 27, 2019.

Animal studies: Not applicable.

DATA AVAILABILITY STATEMENT

Due to the privacy policy, the data of the PHS 2014/15 used in this study are not allowed to be openly accessed. The data will be available upon request to the Hong Kong Department of Health.

REFERENCES

1. Roglic G. WHO Global report on diabetes: a summary. *Int J Noncommun Dis* 2016; 1: 3.
2. Siu AL. Screening for abnormal blood glucose and type 2 diabetes mellitus: US Preventive Services Task Force recommendation statement. *Ann Intern Med* 2015; 163: 861–868.
3. Saeedi P, Petersohn I, Salpea P, *et al.* Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: results from the International Diabetes Federation Diabetes Atlas. *Diabetes Res Clin Pract* 2019; 157: 107843.

4. Plantinga LC, Crews DC, Coresh J, *et al.* Prevalence of chronic kidney disease in US adults with undiagnosed diabetes or prediabetes. *Clin J Am Soc Nephrol* 2010; 5: 673–682.
5. Wang L, Gao P, Zhang M, *et al.* Prevalence and ethnic pattern of diabetes and prediabetes in China in 2013. *JAMA* 2017; 317: 2515–2523.
6. Centers for Disease Control and Prevention. National Diabetes Statistics Report, 2020. Atlanta, GA: Centers for Disease Control and Prevention, US Department of Health and Human Services, 2020; 12–15.
7. Tabák AG, Herder C, Rathmann W, *et al.* Prediabetes: a high-risk state for developing diabetes. *Lancet* 2012; 379: 2279.
8. Gerstein HC, Santaguida P, Raina P, *et al.* Annual incidence and relative risk of diabetes in people with various categories of dysglycemia: a systematic overview and meta-analysis of prospective studies. *Diabetes Res Clin Pract* 2007; 78: 305–312.
9. Barr E, Zimmet PZ, Welborn TA, *et al.* Risk of cardiovascular and all-cause mortality in individuals with diabetes mellitus, impaired fasting glucose, and impaired glucose tolerance: the Australian Diabetes, Obesity, and Lifestyle Study (AusDiab). *Circulation* 2007; 116: 151–157.
10. Echouffo-Tcheugui JB, Simmons RK, Prevost AT, *et al.* Long-term effect of population screening for diabetes on cardiovascular morbidity, self-rated health, and health behavior. *Ann Fam Med* 2015; 13: 149–157.
11. World Health Organization. Screening for Type 2 Diabetes: Report of a World Health Organization and International Diabetes Federation Meeting. Geneva: World Health Organization, 2003.
12. Hong Kong Department of Health. Hong Kong Reference Framework for Diabetes Care for Adults in Primary Care Settings, 2017.
13. Bang H, Edwards AM, Bombback AS, *et al.* Development and validation of a patient self-assessment score for diabetes risk. *Ann Intern Med* 2009; 151: 775–783.
14. Chatterton H, Younger T, Fischer A, *et al.* Risk identification and interventions to prevent type 2 diabetes in adults at high risk: summary of NICE guidance. *BMJ* 2012; 345: e4624.
15. Chen L, Magliano DJ, Balkau B, *et al.* AUSDRISK: an Australian Type 2 Diabetes Risk Assessment Tool based on demographic, lifestyle and simple anthropometric measures. *Med J Aust* 2010; 192: 197–202.
16. Kaczorowski J, Robinson C, Nerenberg K. Development of the CANRISK questionnaire to screen for prediabetes and undiagnosed type 2 diabetes. *Can J Diabetes* 2009; 33: 381–385.
17. Zhang M, Zhang H, Wang C, *et al.* Development and validation of a risk-score model for type 2 diabetes: a cohort study of a rural adult Chinese population. *PLoS One* 2016; 11: e0152054.
18. Zhou X, Qiao Q, Ji L, *et al.* Nonlaboratory-based risk assessment algorithm for undiagnosed type 2 diabetes developed on a nation-wide diabetes survey. *Diabetes Care* 2013; 36: 3944–3952.
19. Woo YC, Gao B, Lee CH, *et al.* Three-component non-invasive risk score for undiagnosed diabetes in Chinese people: development, validation and longitudinal evaluation. *J Diabetes Investig* 2020; 11: 341–348.
20. Baliunas DO, Taylor BJ, Irving H, *et al.* Alcohol as a risk factor for type 2 diabetes: a systematic review and meta-analysis. *Diabetes Care* 2009; 32: 2123–2132.
21. Cappuccio FP, D'Elia L, Strazzullo P, *et al.* Quantity and quality of sleep and incidence of type 2 diabetes: a systematic review and meta-analysis. *Diabetes Care* 2010; 33: 414–420.
22. Department of Health HKSAR Government. Report of Population Health Survey 2014/15, 2017.
23. Collins GS, Reitsma JB, Altman DG, *et al.* Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *J Br Surg* 2015; 102: 148–158.
24. World Health Organization. Use of Glycated Haemoglobin (HbA1c) in Diagnosis of Diabetes Mellitus: Abbreviated Report of A WHO Consultation. Geneva: World Health Organization, 2011.
25. American Diabetes Association. 2. Classification and diagnosis of diabetes: standards of medical care in diabetes – 2019. *Diabetes Care* 2019; 42(Supplement 1): S13–S28.
26. Wong MCS, Sin CKL, Lee JPM. The reference framework for diabetes care in primary care settings. *Hong Kong Med J* 2012; 18: 238.
27. Bush K, Kivlahan DR, McDonnell MB, *et al.* The AUDIT alcohol consumption questions (AUDIT-C): an effective brief screening test for problem drinking. *Arch Intern Med* 1998; 158: 1789–1795.
28. Armstrong T, Bull F. Development of the World Health Organization global physical activity questionnaire (GPAQ). *J Public Health* 2006; 14: 66–70.
29. O'Brien RM. A caution regarding rules of thumb for variance inflation factors. *Qual Quant* 2007; 41: 673–690.
30. Chen T, Guestrin C, editors. Xgboost: A scalable tree boosting system. Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, 2016.
31. Niculescu-Mizil A, Caruana R, editors. Predicting good probabilities with supervised learning. Proceedings of the 22nd international conference on Machine learning, 2005.
32. Lundberg SM, Lee S-I, (eds). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 2017.
33. Kursu MB, Rudnicki WR. Feature selection with the Boruta package. *J Stat Softw* 2010; 36: 1–13.
34. Youden WJ. Index for rating diagnostic tests. *Cancer* 1950; 3: 32–35.
35. Hosmer DW Jr, Lemeshow S, Sturdivant RX. Applied Logistic Regression. Hoboken, NJ: John Wiley & Sons, 2013.

36. Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS One* 2015; 10: e0118432.
37. Van Calster B, McLernon DJ, Van Smeden M, et al. Calibration: the Achilles heel of predictive analytics. *BMC Med* 2019; 17: 1–7.
38. Van Calster B, Nieboer D, Vergouwe Y, et al. A calibration hierarchy for risk models was defined: from utopia to empirical data. *J Clin Epidemiol* 2016; 74: 167–176.
39. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988; 44: 837.
40. Boyd K, Eng KH, Page CD, editors. Area under the precision-recall curve: point estimates and confidence intervals. Joint European conference on machine learning and knowledge discovery in databases. Berlin, Germany: Springer, 2013.
41. Pencina MJ, D'Agostino RB Sr, D'Agostino RB Jr, et al. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med* 2008; 27: 157–172.
42. Kengne AP, Beulens JW, Peelen LM, et al. Non-invasive risk scores for prediction of type 2 diabetes (EPIC-InterAct): a validation of existing models. *Lancet Diabetes Endocrinol* 2014; 2: 19–29.
43. Tsukiyama H, Nagai Y, Matsubara F, et al. Proposed cut-off values of the waist circumference for metabolic syndrome based on visceral fat volume in a Japanese population. *J Diabetes Investig* 2016; 7: 587–593.
44. Ruan Y, Mo M, Joss-Moore L, et al. Increased waist circumference and prevalence of type 2 diabetes and hypertension in Chinese adults: two population-based cross-sectional surveys in Shanghai, China. *BMJ Open* 2013; 3: e003408.
45. WH Organization. Waist Circumference and Waist-hip Ratio: Report of a WHO Expert Consultation. Geneva, 8–11 December 2008: WH Organization, 2011.
46. Feller S, Boeing H, Pischon T. Body mass index, waist circumference, and the risk of type 2 diabetes mellitus: implications for routine clinical practice. *Deutsches Ärzteblatt Int* 2010; 107: 470.
47. Mitsuhashi K, Hashimoto Y, Tanaka M, et al. Combined effect of body mass index and waist-height ratio on incident diabetes; a population based cohort study. *J Clin Biochem Nutr* 2017; 61: 118–122.
48. American Diabetes Association. Waist circumference thresholds provide an accurate and widely applicable method for the discrimination of diabetes. *Diabetes Care* 2007; 30: 3116–3118.
49. Choi S-J, Keam B, Park S-H, et al. Appropriate waist circumference cut-offs to predict diabetes in the Korean population—the Korean genome and epidemiology study. *Circ J* 2010; 74: 1357–1363.
50. Ross R, Neeland IJ, Yamashita S, et al. Waist circumference as a vital sign in clinical practice: a Consensus Statement from the IAS and ICCR Working Group on Visceral Obesity. *Nat Rev Endocrinol* 2020; 16: 177–189.
51. Shan Z, Ma H, Xie M, et al. Sleep duration and risk of type 2 diabetes: a meta-analysis of prospective studies. *Diabetes Care* 2015; 38: 529–537.
52. Honda T, Kuwahara K, Nakagawa T, et al. Leisure-time, occupational, and commuting physical activity and risk of type 2 diabetes in Japanese workers: a cohort study. *BMC Public Health* 2015; 15: 1–9.
53. Villegas R, Shu X-O, Li H, et al. Physical activity and the incidence of type 2 diabetes in the Shanghai women's health study. *Int J Epidemiol* 2006; 35: 1553–1562.
54. Steinbrecher A, Erber E, Grandinetti A, et al. Physical activity and risk of type 2 diabetes among native Hawaiians, Japanese Americans, and Caucasians: the multiethnic cohort. *J Phys Act Health* 2012; 9: 634–641.
55. Aune D, Norat T, Leitzmann M, et al. Physical activity and the risk of type 2 diabetes: a systematic review and dose-response meta-analysis. *Eur J Epidemiol* 2015; 30: 529–542.
56. Noble D, Mathur R, Dent T, et al. Risk models and scores for type 2 diabetes: systematic review. *BMJ* 2011; 343: d7163.
57. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat Mach Intell* 2019; 1: 206–215.

SUPPORTING INFORMATION

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

Figure S1 | Important interactive effects of risk factors on the relative risk of diabetes mellitus and pre-diabetes mellitus by ML modeling ($N = 1238$).

Figure S2 | Interface of software for diabetes mellitus and pre-diabetes mellitus risk assessment.

Figure S3 | The exploration of multicollinearity among BMI, WHR, and waist circumference.

Table S1 | Performance of the risk models based on repeated randomized data splitting.

Table S2 | Association between sleep duration level and risk of new diabetes mellitus and pre-diabetes mellitus ($N = 1238$).