Check for updates

Development and validation of a machine learning-augmented algorithm for diabetes screening in community and primary care settings: A population-based study

XiaoHuan Liu[1,2], Weiyue Zhang[1,2], Qiao Zhang[3], Long Chen[4], TianShu Zeng[1,2], JiaoYue Zhang[1,2], Jie Min[1,2], ShengHua Tian[1,2], Hao Zhang[1,2], Hantao Huang[5], Ping Wang[6], Xiang Hu[1,2]* and LuLu Chen[1,2]*

[1]Department of Endocrinology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, [2]Hubei provincial Clinical Research Center for Diabetes and Metabolic Disorders, Wuhan, China, [3]Department of Cardiovascular Surgery, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, [4]Department of Computer Science and Technology, Tsinghua University, Beijing, China, [5]Yiling Hospital, Yichang, China, [6]Precision Health Program, Department of Radiology, College of Human Medicine, Michigan State University, East Lansing, MI, United States

Background: Opportunely screening for diabetes is crucial to reduce its related morbidity, mortality, and socioeconomic burden. Machine learning (ML) has excellent capability to maximize predictive accuracy. We aim to develop ML-augmented models for diabetes screening in community and primary care settings.

Methods: 8425 participants were involved from a population-based study in Hubei, China since 2011. The dataset was split into a development set and a testing set. Seven different ML algorithms were compared to generate predictive models. Non-laboratory features were employed in the ML model for community settings, and laboratory test features were further introduced in the ML+lab models for primary care. The area under the receiver operating characteristic curve (AUC), area under the precision-recall curve (auPR), and the average detection costs per participant of these models were compared with their counterparts based on the New China Diabetes Risk Score (NCDRS) currently recommended for diabetes screening.

Results: The AUC and auPR of the ML model were 0·697and 0·303 in the testing set, seemingly outperforming those of NCDRS by 10·99% and 64·67%, respectively. The average detection cost of the ML model was 12·81% lower than that of NCDRS with the same sensitivity (0·72). Moreover, the average detection cost of the ML+FPG model is the lowest among the ML+lab models and less than that of the ML model and NCDRS+FPG model.

Conclusion: The ML model and the ML+FPG model achieved higher predictive accuracy and lower detection costs than their counterpart based on NCDRS. Thus, the ML-augmented algorithm is potential to be employed for diabetes screening in community and primary care settings.

Introduction

Diabetes is highly prevalent among adults worldwide and the prevalence has been expanding in developing societies including China and India, consequently increasing the incidence of multiple diabetes complications including cardiovascular disease, retinopathy, kidney disease, neuropathy, blindness, and lower-extremity amputation (1). These complications result in increased morbidity and mortality and impose a heavy economic burden on patients and their health care systems globally (2).

Fortunately, more evidence indicates that early diagnosis and opportune management of diabetes can prevent or delay diabetic complications and dramatically alleviate its related morbidity, mortality, and economic burden (3–5). Diabetes screening is crucial in the early detection and diagnosis of diabetes. Notably, it is estimated that 50·1% of adults (231·9 million) living with diabetes were undiagnosed around the world in 2019, since diabetes usually has a long asymptomatic phase and blood glucose testing is not always available or accessible in communities (1, 6). Additionally, fasting plasma glucose (FPG) is routinely used for diabetes screening but there are a large number of patients with isolated post-load hyperglycemia, which makes a large proportion of patients with missed diagnosis of diabetes (around 38% in our previous study) (7). Moreover, further confirmatory tests such as OGTT (oral glucose tolerance test) or HbA1c cost more and are time-consuming (8). Therefore, it would be extremely useful for timely diagnosis and treatment of diabetes to develop an easy-to-use, very convenient and accessible, and economical screening system with superior sensitivity and specificity to screen out the residents with a high risk of diabetes in the community for further confirmatory test of OGTT, and sift out the individuals likely to have isolated post-load diabetes in primary care for the confirmatory test to detect diabetes.

Risk-based screening for diabetes is currently recommended by ADA (America Diabetes Association) and the Chinese Diabetes Society (CDS) (9, 10). The ADA recommends the use of the scoring table of ADA risk test (ADART), and CDS recommends the use of the New Chinese Diabetes Risk Score (NCDRS) to screen out high-risk individuals for the further confirmative test. These two scoring tables developed by Logistic regression (LR) based on easy-to-accessible non-laboratory characteristics are of great help in clinical practice for diabetes screening (8, 11). Notably, there is increasing evidence indicating that machine learning (ML) methods are able to predict relationships between input and output by extracting information from a larger number of complex variables and have shown accurate predictive ability (12, 13). It has been increasingly introduced in health service research and has shown a better level of prediction than traditional statistical approaches in several domains (14–16). Recently, it is reported by Vangeepuram, N., et al. that Some ML-based classifiers derived from the National Health and Nutritional Examination Survey (NHANES) dataset in the United States performed comparably to or better than the screening guideline in identifying preDM/DM youth (17). Herein, we hypothesized that the ML-based algorithm has the potential to develop predictive models for diabetes screening and be helpful to screen out people with a high risk of diabetes in the community and individuals likely to have isolated post-load diabetes in primary care more accurately and conveniently for further confirmation and diagnosis opportunely, which would outperform standard risk-scoring algorithms for screening diabetes and increase the cost-effectiveness of detection by reducing the false-positive screening rate.

In this study, we aimed to develop and validate a ML-augmented algorithm for diabetes screening in community and primary care settings with data from a population-based study in China.

Methods

Study design, participants, and data collection

The data analyzed in the present investigation were obtained from the Hubei Yiling center of the Risk Evaluation of cAncers

in Chinese diabeTic Individuals: a lONgitudinal (REACTION) study performed from October 2011. Detailed information and the study design of the REACTION study were described previously (18). A total of 10184 eligible subjects were enrolled in this study. Data collection was performed by the trained staff and a questionnaire was completed as described previously for gathering information on demographic characteristics, data on lifestyle, and medical history (7). Weight, height, waist circumference (WC), hip circumference (HC), blood pressure, and resting pulse rate (RPR) were measured according to standard protocols. Body mass index (BMI), waist-to-hip ratio (WHR), and waist-to-height ratio (WHtR) were calculated as described previously (19). Participants without a self-reported history of diabetes were provided with a standard 75 g glucose solution, and blood sampling was conducted at 0 and 2 h after administration. Plasma glucose was measured using the glucose oxidase method. HbA1c was tested using finger capillary whole blood by high-performance liquid chromatography. Participants who had been diagnosed with cancers or diabetes before or using antidiabetic agents, or whose data on FPG, 2hPG or HbA1c were missing were excluded and 8425 participants were included in this analysis. The data collected were analyzed and the flow chart of this study was shown in Figure 1.

The present study complies with the Declaration of Helsinki, and all procedures were approved by the Ethics Committee of Tongji Medical College, Huazhong University of Science and Technology. Written, informed consent was obtained from all the participants.

Outcome

Diabetes was diagnosed based on measurement of fasting blood glucose level, or oral glucose tolerance testing as recommended by CDS when the study was performed, which included the following: a FPG level of ≥7·0mmol/l, or OGTT-2h post-load plasma glucose (2hPG) level of ≥11·1 mmol/l.

Predictors and data processing

150 non-laboratory (demographics and anthropometric) and three laboratory features (FPG, 2hPG, HbA1c), were included in the analysis after excluding features with more than 20% missing data as described previously (20, 21). MissForest was used to impute missing values for the features with less than 20% missing data as described previously (22). Data were divided into a development set (60%) and a testing set (40%) and the imputation was trained in the development and applied to the testing set to avoid data leakage. A complete case sensitivity analysis was performed to compare the difference before and after imputation of missing data (Supplementary
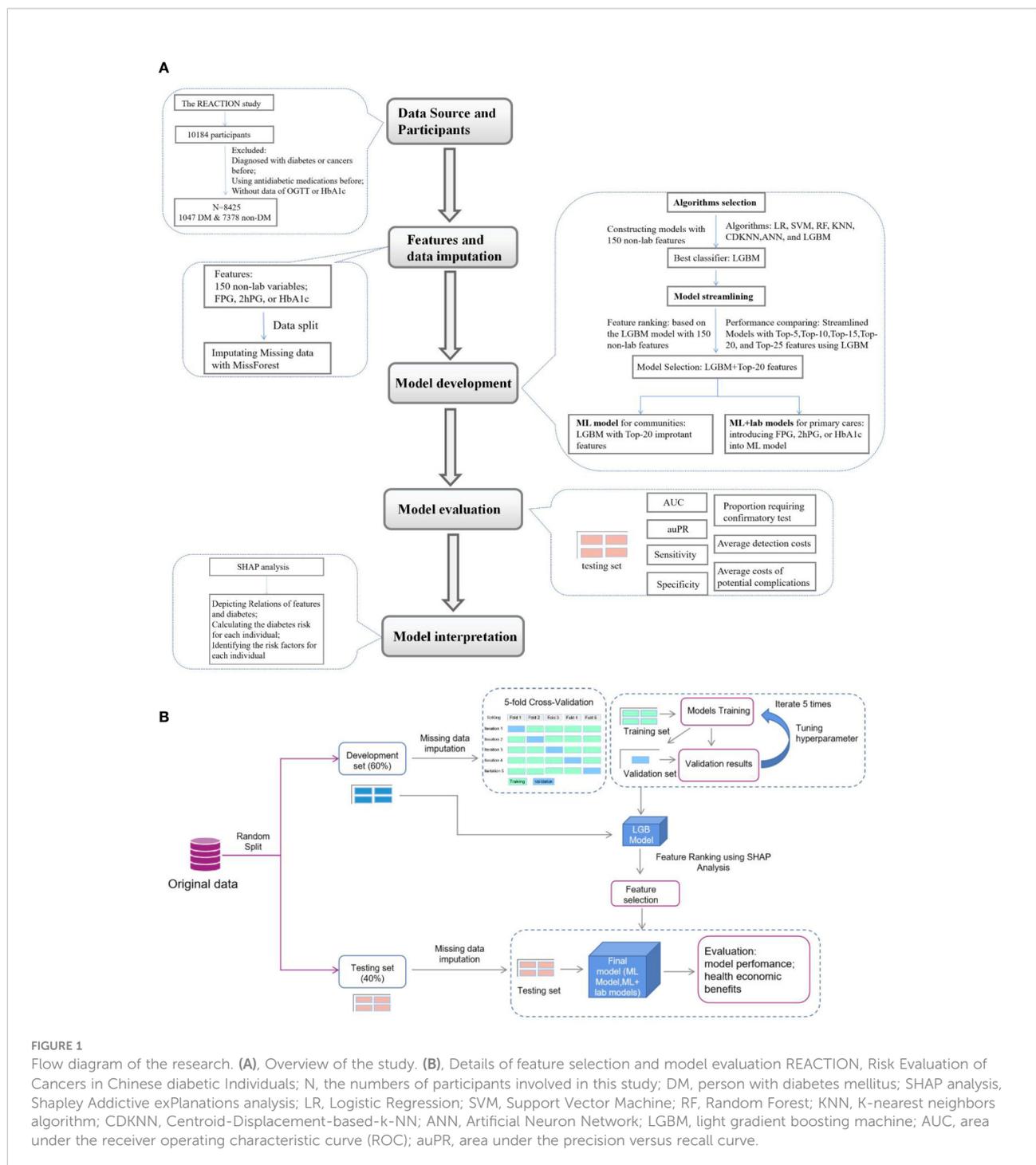
Tables 1, 2). The development set was used for training and validating models. The testing set was blind to the training, hyperparameter tuning, and feature selection, and were only used to evaluate the performance and health economic benefits of models.

Models development by machine learning for the community and primary care

Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF), K-nearest neighbors algorithm (KNN), Centroid-Displacement-based-k-NN (CDKNN) (23), Artificial Neuron Network (ANN), and Light Gradient Boost Machine (LGBM) were preliminarily tested as the classification algorithms to develop the predictive models for diabetes screening with 150 non-laboratory features. TPE (Tree of Parzen Estimators) was used to tune hyperparameters and improve model prediction capability (24). The hyperparameters were reported in Supplementary Table 3. For internal validation, 5-fold cross-validation was used as reported previously (25–27). In detail, we randomly and equally split the development set into 5 folds using the random.shuffle() function of the Numpy library in Python, and for each iteration, we employed four folds to train the models, and the remaining fold to validate the models independently. The algorithm that had the best predictive ability was selected to develop the predictive models in this study.

Shapley Additive Explanations (SHAP) analysis (28) was employed to interpret the results and analyze the importance of (the contribution of a feature value to the difference between the actual prediction and the mean prediction is the estimated Shapley value) of 150 non-laboratory features of the model with the best performing algorithm (LGBM). Five streamlined models were developed using 5, 10, 15, 20, 25 features of top-ranked importance among all the non-laboratory features with LGBM to simplify the predictive model for practice. The streamlined model with the best predictive power and the least number of features was selected and noted as the ML model for diabetes screening in community care.

A large number of individuals with seemingly normal levels of FPG (FPG<7.0mmol/L) or 2hPG (2hPG<11.0mmol/l) are potentially diabetes patients and too many patients would not be diagnosed opportunely using only one testing mentioned above. However, these seemingly normal levels of FPG, 2hPG, or HbA1c are likely to be useful in the prediction of diabetes. Thus, we tried to develop the ML+lab models by introducing one test (seemingly normal FPG, 2hPG, or HbA1c) in the hope of getting an efficient, cost-effective, and convenient screening model in primary care to decrease the missed diagnoses of diabetes and related costs (Figure 2). That is, the ML+FPG, ML+2hPG or ML +HbA1c models predicted isolated post-load diabetes (identified

FIGURE 1

Flow diagram of the research. **(A)**, Overview of the study. **(B)**, Details of feature selection and model evaluation REACTION, Risk Evaluation of Cancers in Chinese diabetic Individuals; N, the numbers of participants involved in this study; DM, person with diabetes mellitus; SHAP analysis, Shapley Addictive exPlanations analysis; LR, Logistic Regression; SVM, Support Vector Machine; RF, Random Forest; KNN, K-nearest neighbors algorithm; CDKNN, Centroid-Displacement-based-k-NN; ANN, Artificial Neuron Network; LGBM, light gradient boosting machine; AUC, area under the receiver operating characteristic curve (ROC); auPR, area under the precision versus recall curve.

using 2hPG) by introducing seemingly normal FPG levels, and isolated fasting diabetes (identified using FPG) by seemingly normal 2hPG levels and diabetes (identified using FPG and 2hPG) by normal HbA1c, respectively.

The ML model was compared with the ADART and NCDRS. The ML+lab models were compared with the ML model and with the corresponding NCDRS+lab models developed by combining FPG, 2hPG, or HbA1c and the NCDRS.

Model performance evaluation

We used the area under the receiver operating characteristic curve (AUC), the area under the precision-recall curve (auPR), sensitivity, specificity, and precision to evaluate model predictive ability. True positive (TP), true negative (TN), false positive (FP), false negative (FN) cases, and negative cases were calculated for further analysis. Sensitivity (recall), a measure of
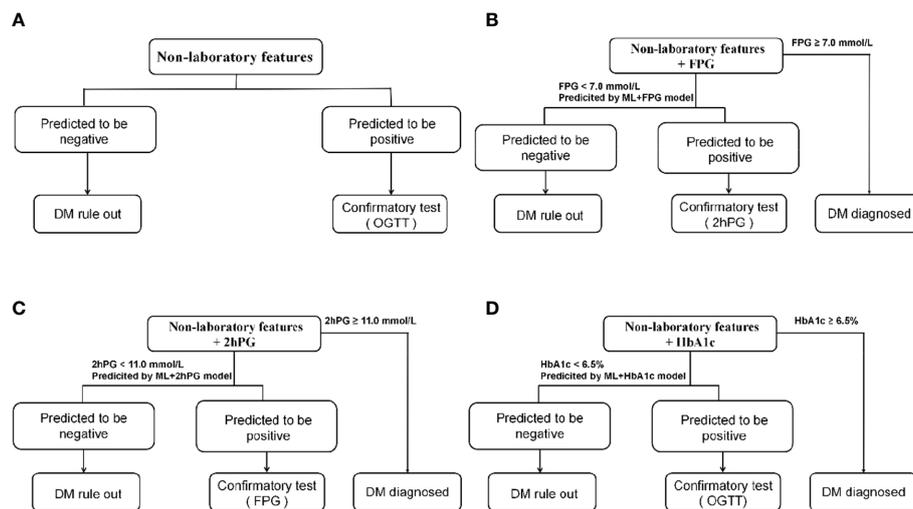
FIGURE 2
Diabetes detection procedures of models. **(A)**, Detection procedures of ML model and NCDRS. **(B)**, Detection procedures of ML+FPG model and NCDRS+FPG model. **(C)**, Detection procedures of ML+2hPG model and NCDRS+2hPG model. **(D)**, Detection procedures of ML+HbA1c model and NCDRS+HbA1c model. OGTT in a and d included fasting plasma glucose and 2h post-load plasma glucose. NCDRS, New Chinese Diabetes Risk Score; DM, diabetes mellitus; OGTT, oral glucose tolerance test; FPG, fasting plasma glucose; 2hPG, OGTT-2h post-load plasma glucose, HbA1c, glycated hemoglobin A1c.

the ability of the model to identify diabetes, was defined as TP/(TP+FN). Specificity, the proportion predicted to be negative among the non-diabetes population, was calculated by TN/(FP+TN). Precision was defined as TP/(TP+FP). AUC was calculated from the curve of sensitivity against 1-specificity and auPR was calculated from the precision-recall curve.

Health economic evaluation

Individuals at high risk of diabetes are recommended to perform the confirmatory test in clinical practice according to guidelines (9, 10). Therefore, we attempted to identify the risk of individuals with the ML model, ML+lab models, NCDRS, or NCDRS+lab models in the present study, to screen out those with high risk for the further confirmatory test. The individuals requiring confirmatory test in this study were those who were predicted to be positive and screening tests of FPG < 7·0 mmol/l or 2hPG < 11·0 mmol/l (if available) in the testing set. The proportion requiring confirmatory test was calculated by dividing the numbers of individuals requiring confirmatory test by the number of participants in the testing set (Figure 2; Supplementary Table 4).

The costs associated with screening tests (FPG, 2hPG, or HbA1c in the ML/NCDRS+lab models) and confirmatory tests were made up of medical costs and non-medical costs (e.g transportation costs). Medical costs (CNY) for FPG, 2hPG, OGTT, and HbA1c were 9·89, 23·56, 33·45, and 84·16, respectively as described previously (7, 29), and non-medical costs (CNY) calculated based on the report described previously

were 8·3, 27·5, 27·5, and 8·3 for FPG, 2hPG, OGTT, and HbA1c, respectively (30). A further confirmatory test is required after the diabetes screening, a process known as diabetes detection. The average detection costs per participant of using predictive models as a screening strategy were calculated by dividing the sum of the costs associated with screening and confirmatory test by the number of participants in the testing set (Supplementary Table 4).

In view that early diagnosis and opportune management of diabetes can prevent or delay diabetic complications, in the present study we tried to assess the potential complication costs, i.e. costs of false negative were assessed by estimating the increased costs from complications not prevented by the lack of timely and appropriate intervention for diabetes (31). The potential complication costs per individual with diabetes per year in China were estimated after adjusting for medical cost differences based on a previous study (31), ranging from (CNY) 341 to 567, 1302 to 2555, 2802 to 5611, 4428 to 8212, and 5258 to 9132 over 5, 10, 15, 20, and 25 years, respectively. The average costs of potential complication per participant of using predictive models as a screening strategy were calculated by dividing the sum of potential complication costs of all cases of FN by the number of participants in the testing set (Supplementary Table 5) (7).

Model interpretation

SHAP analysis was employed for the interpretation of the ML model. In detail, SHAP values were calculated for the top-10 features and converted to relative risk (RR) of diabetes to explore

the relationships between features and diabetes risk as reported in previous studies (32). In addition, two cases of TP were selected randomly as examples to demonstrate practically how the ML model works. In the demonstration, their respective key predictors were classified respectively, and the importance of their respective predictors was assessed by calculating SHAP values, and predictive risk of diabetes was evaluated individually by summing their SHAP values.

Statistical analysis

Data of participants were presented as medians (IQRs) for continuous variables and numbers (proportions) for categorical variables. These data were tested for normality using the Kolmogorov-Smirnov test. The Kruskal-Wallis test was used to compare continuous variables (skewed variables) and the chi-square test was used to compare categorical variables and the Delong test was used to compare AUC. A 2-sided P value <0·05 was considered statistically significant. Data were analyzed using Python 3·7 and SPSS 20·0. MissForest, LR, SVM, and RF were implemented with the ML library "sklearn" (33), and the ANN was implemented with PyTorch. LGBM referred to https://lightgbm.readthedocs.io/. SHAP analysis referred to http://github.com/slundberg/shap (28).

Results

Characteristics of participants

In the present study, 1047 (12·4%) were diagnosed with diabetes. Principally, the age, resting pulse rate, blood pressures, weight, BMI, WC, WHR, WHtR, FPG, 2hPG, and HbA1c were higher, and the amount of physical activity and alcohol consumption was lower (P<0·05) in participants with diabetes compared to individuals without diabetes, while no significant differences were observed in sex and numbers of smoker (P>0·05) (Table 1).

Development and interpretation of the ML model for community care

Our results indicated that the predictive model developed by LGBM for diabetes screening seemingly had the best performance, with the highest auPR of 0·319 [95%CI, 0·267-0·386] and AUC of 0·691[95%CI, 0·641-0·733] (Figure 3A; Supplementary Table 6), among the models created by the seven ML algorithms (LR, KNN, CDKNN, SVM, LR, ANN, LGBM). To simplify the used features in the model, five streamlined models were developed with LGBM using the top-5, top-10, top-15, top-20, and top-25 features, respectively,

according to the feature importance. Among these models, the AUC (0·699 [95%CI, 0·663-0·736]) of the streamlined model with top-20 features was highest (Figure 3B). While the auPR of the streamlined model with top-20 features (0·301[95%CI, 0·220, 0·390]) was closest to that of the model with all features. The model with top-20 features seemed a fairly convenient and accurate model and was adopted for the subsequent analysis, noted as the ML model in the present study.

The top-20 important features of the ML model included resting pulse rate (RPR), Age, systolic pressure (SBP), waist circumference (WC), work status, WHR, etc (Figure 3C). The most important predictors associated with the predictive power of the model were RPR, age, and SBP, which were positively associated with RR of diabetes (when RPR was higher than 78, age was higher than 52 years, or SBP exceeded 135, the RR of diabetes >1) (Figures 3D–F). The RRs of diabetes for other top-10 important features were shown in Supplementary Figure 1. In addition, SHAP analysis performed by taking two cases, for example, indicated that individualized important predictors were identified for case1(including RPR, SBP, age, etc) and case2 (including WHR, WC, and age, etc). The magnitudes of these predictors were also assessed (Figures 3G, H).

Comparison of the predictive ability between ML and ML+lab models with their corresponding counterparts

The 150 features of our predictive model included all the features used in ADART and NCDRS, and the top-20 features of the ML model comprised most of the features used in ADART (4 of 7) and NCDRS (4 of 6). In addition, our results indicated that 15 features, which were not included in ADART and NCDRS, were also very important in the ML model (Figure 4A).

The AUC was 0·697 in the ML model in the testing set, which was higher than that in the ADART (0·608) and NCDRS (0·628). The auPR of the ML model (0·303) was also higher than that of the ADART (0·168) and NCDRS (0·184) (Figures 4B, C, Supplementary Table 7). Remarkably, the analysis in our study indicated that the NCDRS had higher AUC and auPR compared with the ADART. Therefore, we chose to perform further comparisons between our models and NCDRS on sensitivity, specificity and health economic evaluation. Our analysis indicated that the sensitivity and specificity of the NCDRS were 0·72 and 0·45 respectively using the cutoff recommended by CDS, which is considered to be optimal in the Chinese population (10). The ML model had higher sensitivity (0·80 vs 0·72) when it had the same specificity as NCDRS (0·45). Likewise, The ML model had higher specificity (0·53 vs 0·45) when it had the same sensitivity as NCDRS (0·72).

Our results indicated that 41% of diabetes patients have seemingly normal FPG (FPG<7·0mmol/l), 28% of them have seemingly normal 2hPG (2hPG<11·0mmol/l), and only 31% of

TABLE 1 Characteristics of participants.

	Total (n=8425)	Non-DM (n=7378)	DM (n=1047)	P
Age, years[a]	54·0 (47·0-60·0)	54·0 (47·0-60·0)	57·0 (50·0-63·0)	0·000
Sex, n (%)[b]				0·090
Male	2989 (35·5)	2593 (35·1)	396 (37·8)	
Female	5436 (64·5)	4785 (64·9)	651 (62·2)	
Smoke, n (%)[b]				
Current smoker	1621 (19·2)	1427 (19·3)	194 (18·5)	0·789
Past smoker	353 (4·2)	307 (4·2)	46 (4·4)	
Never smoker	6451 (76·6)	5644 (76·5)	807 (77·1)	
Drink, n (%)[b]				0·006
Current drinker	1088 (12·9)	921 (12·5)	167 (16·0)	
Past drinker	203 (2·4)	176 (2·4)	27 (2·6)	
Never drinker	7134 (84·7)	6281 (85·1)	853 (81·5)	
Physical activity, n (%)[b]				0·044
inactive	7595 (90·1)	6633 (89·9)	962 (91·9)	
active	830 (9·9)	745 (10·1)	85 (8·1)	
SBP, mmHg[a]	135·0 (122·5-150·7)	134·0 (122·0-149·3)	141·0 (128·0-157·5)	0·000
DBP, mmHg[a]	79·0 (71·0-87·0)	78·5 (71·0-86·7)	81·7 (74·0-90·3)	0·000
RPR, bpm[a]	80·0 (72·3-88·0)	79·0 (72·0-87·0)	85·0 (77·0-95·3)	0·000
Weight, kg[a]	57·0 (50·7-63·0)	56·5 (50·5-63·0)	58·5 (52·0-65·0)	0·000
Height, m[a]	1·56 (1·52-1·62)	1·56 (1·52-1·62)	1·56 (1·51-1·62)	0·306
WC, cm[a]	78·0 (72·0-84·0)	77·5 (71·0-84·0)	80·3 (74·0-87·0)	0·000
HC, cm[a]	90·0 (86·0-95·0)	90·0 (86·0-95·0)	91·0 (87·0-96·0)	0·005
BMI, kg/m2[a]	23·1 (21·0-25·4)	23·0 (20·9-25·2)	23·9 (21·6-26·3)	0·000
WHR, cm/cm[a]	0·86 (0·81-0·90)	0·86 (0·81-0·90)	0·89 (0·83-0·93)	0·000
WHtR, cm/m[a]	0·50 (0·46-0·54)	0·49 (0·46-0·54)	0·52 (0·47-0·55)	0·000
FPG, mmol/l[a]	5·7 (5·3-6·1)	5·6 (5·3-6·0)	7·1 (6·2-8·0)	0·000
2hPG, mmol/l[a]	6·9 (5·8-8·3)	6·6 (5·7-7·8)	12·2 (10·3-14·9)	0·000
HbA1c, %[a]	5·5 (5·2-5·8)	5·5 (5·2-5·8)	6·0 (5·6-6·6)	0·000

[a]Data (continuous variables non-normally distributed) are expressed as medians (interquartile range).
[b]Data (categorical variables) are expressed as N (%).
DM, diabetes mellitus; SD, standard deviation; n (%), number of participants and percentage over the total number of participants; SBP, systolic blood pressure; DBP, diastolic blood pressure; RPR, resting pulse rate; WC, waist circumference; HC, hip circumference; BMI, body mass index; WHR, waist-to-hip ratio; WHtR, waist-to-height ratio; FPG, fasting plasma glucose; 2hPG, OGTT-2h post-load plasma glucose; HbA1c, glycated hemoglobin A1c.

them have a combination of increased FPG (FPG ≥7·0mmol/l) and 2hPG (2hPG ≥11·0mmol/l) levels, which are both higher than the diagnostic criteria (Supplementary Figure 2), implying that using one testing of FPG or 2hPG only would lead to numerous missed diagnosis of diabetes.

The results showed that the AUC of the ML+FPG model was 0·740, which was higher by 8.5% compared with the ML model (0·682) in individuals with seemingly normal FPG levels (P<0·05). Likewise, the ML+HbA1c model had a 8.0% increase compared to the ML model in individuals with seemingly normal HbA1c levels (P<0·05). While the AUC of ML+2hPG model had no statistical difference with that of the ML model in individuals with seemingly normal 2hPG levels (P>0·05) (Figures 4D–F).

Moreover, the AUC of the ML+FPG model was higher than that of the NCDRS+FPG model in seemingly normal FPG individuals (P<0·05). Likewise, the AUC were higher in the ML+2hPG and ML

+HbA1C models compared with their corresponding counterparts based on NCDRS (Supplementary Figure 3).

Comparison of the health economic costs between ML and ML+lab models with their corresponding counterparts

Our results indicated that the proportion requiring confirmatory test and the average detection costs per participant using the ML model were lower than those using NCDRS at any sensitivity. The proportion requiring confirmatory test was 57·96% and 50·07% in NCDRS and our ML model when they had the same sensitivity of 0·72, and consequently, their average detection costs per participant were ¥34.98 and ¥30·41 respectively in the analysis of the present study. These results suggested that the proportion requiring
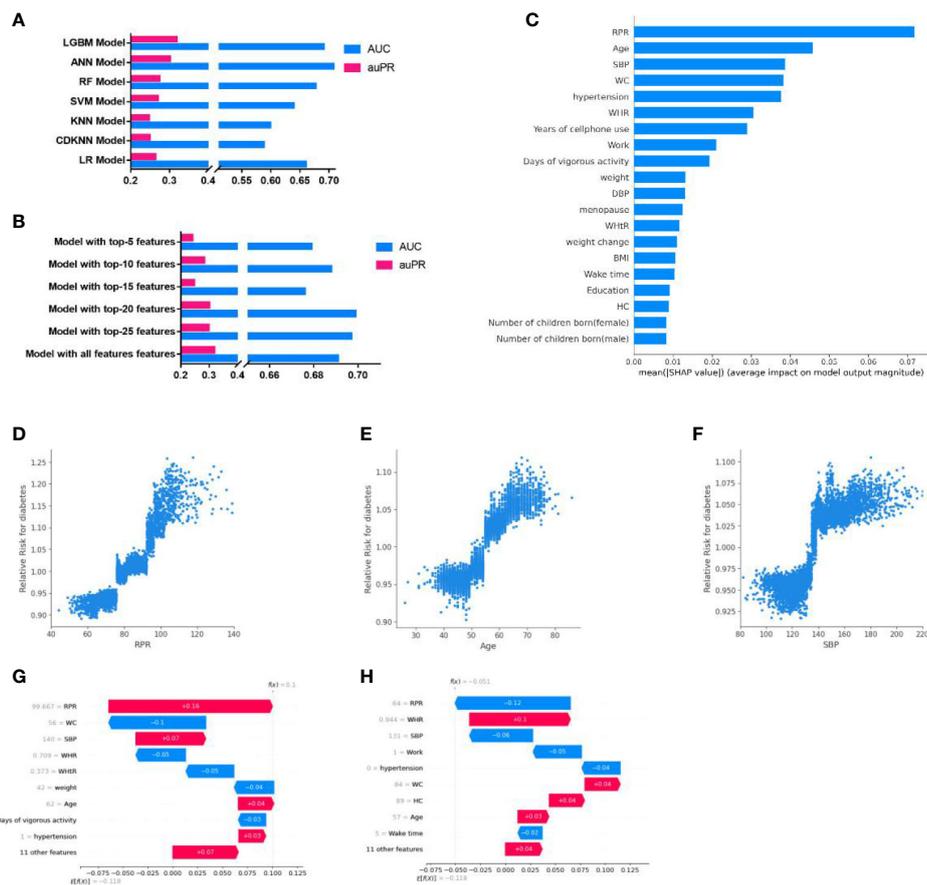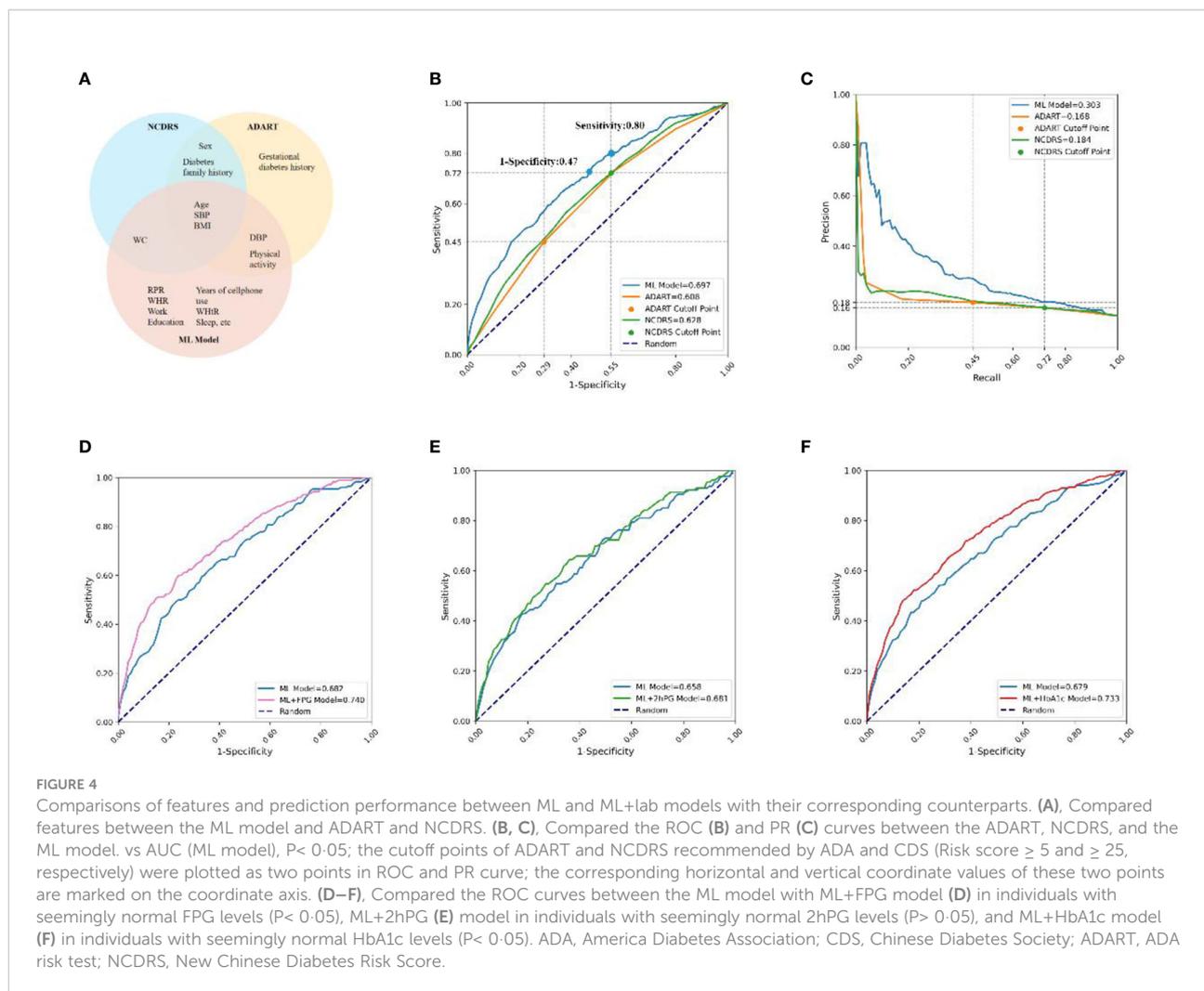
FIGURE 3

Development and interpretation of the ML model. **(A)**, Compared the AUC and auPR among the models with seven algorithms. **(B)**, Compared the AUC and auPR among streamlined models and the model with all features. **(C)**, Importance of the features in the ML model. The input features on the y-axis are ordered by descending importance and the values on the x-axis indicate the mean impact of each feature on model output magnitude based on SHAP analysis. **(D−F)**, The relative risk of diabetes of RPR **(D)**, Age **(E)**, and SBP **(F)**. Each point represented the predicted relative diabetes risk of each individual. The relative risk value above 1·0 for specific features represent an increased risk of diabetes. **(G-H)**, Personalized risk prediction for two cases from the validation set of the ML model. The y-axes represent the input features ordered by descending importance. f(x) is the personalized model output for a participant. If f(x) is larger than e[f(x)], the patient has a higher risk of diabetes relative to the background population. Each arrow represents how a specific feature increases (red) or decreases (blue) the participant's risk for diabetes. LGBM, light gradient boosting machine; ANN, Artificial Neuron Network; RF, Random Forest; SVM, Support Vector Machine; LR, Logistic Regression; KNN, K-nearest neighbors algorithm; CDKNN, Centroid-Displacement-based-k-NN; RPR, resting pulse rate; WHR, waist-to-hip ratio; SBP, systolic blood pressure; BMI, body mass index; WC, waist circumference; WHtR, waist-to-height ratio; DBP, diastolic blood pressure; HC, hip circumference.

confirmatory test and the average detection costs per participant were both lower by 12.81% in our ML model than those in the NCDRS (Figures 5A, F, B). In addition, the ML model had higher sensitivity (0·80 vs 0·72) when it had the same average detection costs as NCDRS (¥34.98) and at any average detection costs (Figure 5C). Furthermore, the average costs of potential complications per person per year of the ML model (ranged from ¥78.16 to ¥156.51 for 15 years) decreased by 19.7% compared with the NCDRS (ranged from ¥97.28 to ¥194.8 for 15 years) when it had the same average detection costs as

NCDRS (¥34.98) (Figure 5D). Likewise, the ML model had a lower average costs of potential complications for 15 years per participant than the NCDRS at any average detection costs (Figure 5E).

The average detection costs per participant were lower in the ML+FPG model and were higher in the ML+2hPG and ML+HbA1c models compared with the ML model (Figure 5F). The sensitivity was higher and the average costs of potential complications per participant were lower in the ML+FPG compared with the ML model. However, the sensitivity was

FIGURE 4

Comparisons of features and prediction performance between ML and ML+lab models with their corresponding counterparts. **(A)**, Compared features between the ML model and ADART and NCDRS. **(B, C)**, Compared the ROC **(B)** and PR **(C)** curves between the ADART, NCDRS, and the ML model. vs AUC (ML model), P< 0·05; the cutoff points of ADART and NCDRS recommended by ADA and CDS (Risk score $\geq$ 5 and $\geq$ 25, respectively) were plotted as two points in ROC and PR curve; the corresponding horizontal and vertical coordinate values of these two points are marked on the coordinate axis. **(D−F)**, Compared the ROC curves between the ML model with ML+FPG model **(D)** in individuals with seemingly normal FPG levels (P< 0·05), ML+2hPG **(E)** model in individuals with seemingly normal 2hPG levels (P> 0·05), and ML+HbA1c model **(F)** in individuals with seemingly normal HbA1c levels (P< 0·05). ADA, America Diabetes Association; CDS, Chinese Diabetes Society; ADART, ADA risk test; NCDRS, New Chinese Diabetes Risk Score.

lower and average costs of potential complications for 15 years per participant were higher in ML+2hPG and ML+HbA1c models compared with the ML model (Figures 5G, H).

Moreover, the average detection costs per participant decreased in the ML+FPG model compared with the NCDRS+FPG model. The sensitivity was higher and the average costs of potential complications per participant were lower in the ML+FPG model than in the NCDRS+FPG model. Likewise, the sensitivity was higher, and the average detection costs per participant as well as average complication costs per participant were lower in the ML+2hPG and ML+HbA1C models compared with their corresponding counterparts based on NCDRS (Supplementary Figure 4).

Discussion

In the present study, we attempted to develop an easy-to-use ML-augmented prediction model for diabetes screening using population-based data from China. Our analysis indicated that the ML model developed with non-laboratory features for community care had superior predictive accuracy, and could lower average detection costs per participant by 12.81% compared with the NCDRS even if it had the same sensitivity as NCDRS. Additionally, the ML+lab models which were developed for primary care by adding laboratory tests to the ML model, had even better predictive accuracy. Remarkably, the ML+FPG model had considerable accuracy and lower average detection costs compared with the ML model.

In recent years, there is a growing body of evidence indicating that ML seems very promising in risk predictions for disease (32, 34, 35). In this study, our ML model had higher sensitivity, specificity, AUC, and auPR compared to the NCDRS, implicating that our ML model seemed to have superior predictive accuracy. As a data-driven method, ML is widely considered to be able to detect complex nonlinear relationships and probable interactions between variables and outcomes (36). Additionally, ML is capable of mining more information from
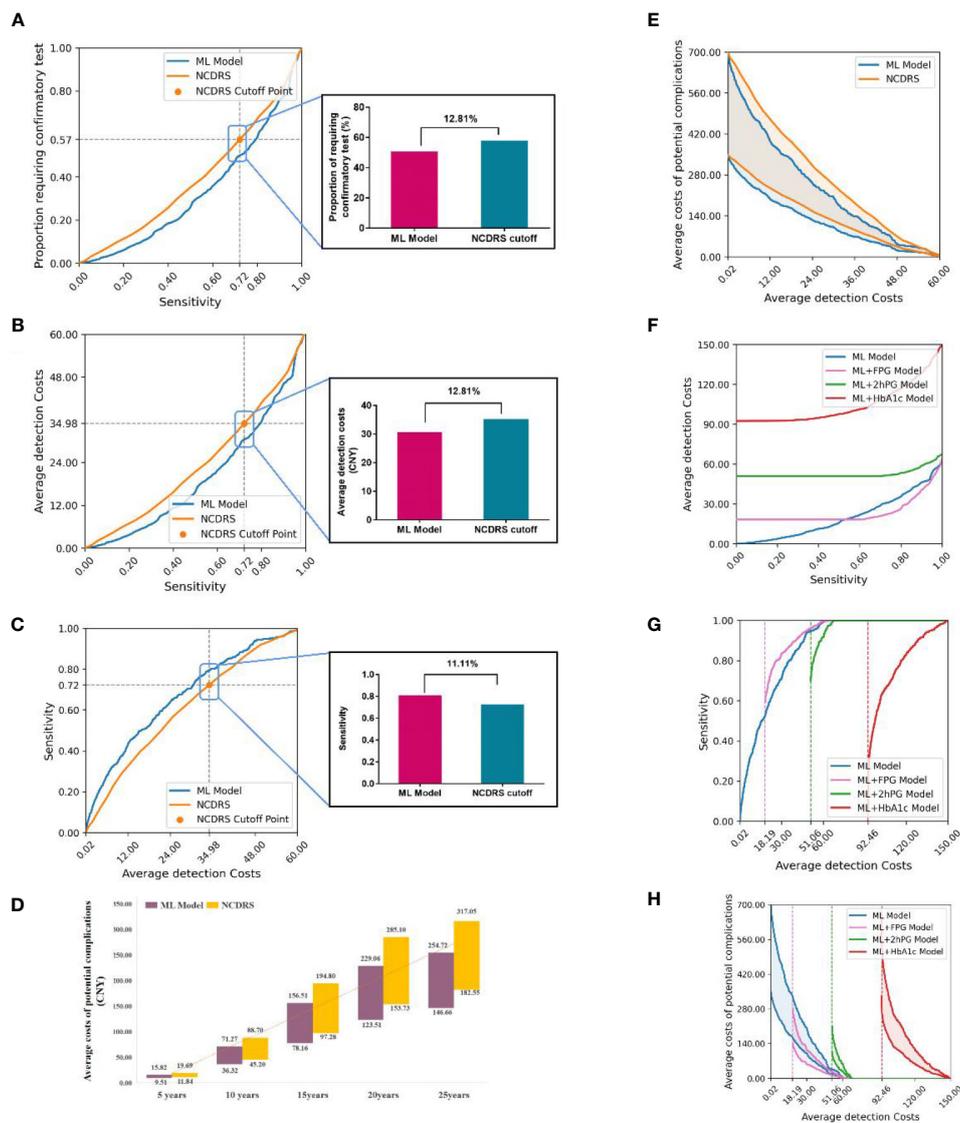
FIGURE 5

Comparison the health economic costs between ML and ML+lab models with their corresponding counterparts. **(A-B)**, Compared the proportion requiring confirmatory test **(A)** and average detection costs **(B)** between the ML model and NCDRS at a different level of sensitivity. **(C)**, Compared the sensitivity between the ML model and NCDRS at a different level of average detection costs. **(D)**, Compared the average costs of potential complications in the next 5-25 years between the ML model and NCDRS when the average detection costs was ¥34·98 (When using the cutoff point of the CDS guideline, the average detection costs of NCDRS were ¥34·98). **(E)**, Compared the average costs of potential complications between the ML model and NCDRS at a different level of average detection costs. **(F)**, Compared the average detection costs at a different level of sensitivity between the ML model with ML+lab models. g-h, Compared the sensitivity **(G)** and the average costs of potential complications for 15 years **(H)** at a different level of average detection costs between the ML model with ML+lab models. The number on the horizontal line **(A-C)** represents the relative reduction of the proportion requiring confirmatory test, average detection costs, and increase of sensitivity of the ML model compared to the NCDRS. The numbers in the bar chart **(D)** represent the range of average costs of potential complications. The dotted lines in **(G, H)** represent the basic average detection costs of the ML+lab models which are the costs of screening test of these models. Shaded areas represent the range of average costs of potential complications for 15 years.

big data compared to Logistic Regression and consequently, more predictors could be handled in predictive models developed by ML (37). Remarkably, there is a rich library of available machine learning methods used for developing models to deal with a specific problem, and it is crucial to select an appropriate ML method to improve the performance of the models (4). Recent studies indicate that diabetes prediction models developed based on images, electronic health records, or structured data obtained from their societies, using machine learning algorithms such as Decision Tree, Naive Bayes, SVM,

ANN, etc., achieve superior performance and demonstrate their potential to be helpful for diabetes screening (38, 39). We employed seven ML algorithms for diabetes screening using data from our population-based study in China, including LGBM, ANN, SVM, RF, KNN, CDKNN and LR, which are reported to have good performances in developing predictive models with high accuracy in recent studies (40–44). Our results indicated that the LGBM model had the best predictive accuracy among the models developed with the five algorithms in our societies. It is considered that LGBM acts as the state of the art in developing predictive models for tabular data (32), and a very recent study indicates that Gradient Boost Machine (GBM) performed better than logistic regression (LR), classification, and regression tree (CART), artificial neural networks (ANN), support vector machine (SVM) and random forest (RF) (45), which might be important explanations of our results that the ML model seemed to have higher predictive accuracy compared with NCDRS. Noteworthy, we tested the ensemble algorithm that combined the classifiers of LR, RF, ANN, and LGBM with the same features as the ML model for predicting diabetes. The results indicated that the ensemble algorithm had a slightly higher AUC than LGBM, while there was no statistically significant difference (Supplementary Figure 5, P>0·05). New algorithms for designing accurate and effective models for diabetes screening still need further investigation in the future.

Notably, we analyzed the performance of our models with methods recommended in the current guideline for screening diabetes in adults and the results indicated that our models seemed to perform better. Similarly, a very recent study by Vangeepuram, N., et al. indicated that some ML-based classifiers derived from the NHANES dataset in the United States performed comparably to or better than the screening guideline in identifying preDM/DM youth, which is another important evidence that ML model seemed to have better performance compared with recommendations of the current guidelines (17). In addition, the diabetes prediction models constructed by Binh P.Nguyen,et al and Wei et,al achieved good prediction performance using 1321 features (including various laboratory tests) and environmental chemical exposures that are not routinely tested in daily life, respectively (46, 47). Our ML model used 20 features that are easily available in daily life and clinical practice in order to make the prediction model more convenient for use in the community and primary care settings. Additionally, we evaluated the proportion requiring confirmatory test and the average detection costs to assess the health economic benefits of the ML models in practice, which is a new attempt compared to previous ML diabetes prediction studies (4, 17, 45, 47–53).

Our analysis demonstrated that the ML model could save the average detection costs per participant by 12.81% compared with NCDRS, without sacrificing sensitivity. The cost saving was mainly attributed to the higher predictive accuracy, which consequently enhanced specificity and decreased the false-positive rate in the ML model, resulting in a considerable decline in the number of confirmatory test and related costs (54). In addition, the ML model could lower the average costs of potential complications per participant by 19.7% compared with NCDRS in this study, which might be ascribed to the increase in predictive accuracy and sensitivity, as well as the decreases in misdiagnosis in the ML model compared with NCDRS, lessening the incidence of complications and corresponding costs for the lack of timely detection and intervention of diabetes.

Remarkably, the features used in our ML model are readily obtainable information, enabling residents to estimate their risk of diabetes in communities without seeing their doctors for any medical examinations or laboratory tests. Most of the variables (RPR, WC, sleep-associated issues, etc) in the ML model were routinely collected in clinical practice in China, and those variables involved in the ML models which are not routinely collected by medical practitioners (education, work status, etc) can be quickly obtained by easy-to-use questionnaires, which are different from those variables in diabetes predictive models including tongue features and environmental chemical exposure developed in very recent studies (47, 51, 52), and might be a little bit more convenient in practice. Thus, the ML model is potentially beneficial for diabetes screening in communities and primary care settings with greater convenience and accessibility, as well as higher predictive accuracy and less cost as mentioned above. Moreover, the ML model could be hopefully further developed and presented in open and accessible web pages to make it easier and more available for residents to evaluate the risk of diabetes in communities, and the information increasingly inputted in the ML model for risk prediction is very helpful to improve the performance of the model in turn, which is an advantage of ML (28).

FPG, 2hPG, or/and HbA1c are important screening and diagnostic tests of diabetes (9). It should be noted that all these tests are not usually employed for detecting diabetes in the same individuals at the same visit in clinical practice and one testing of them only may lead to underdiagnosis of diabetes (9). Actually, there are a large number of diabetes with seemingly normal FPG or 2hPG levels since the concordance between the FPG and 2hPG tests is imperfect (55), which is also observed in our present investigation. It is reported that seemingly normal FPG, 2hPG, and HbA1c were important predictors of future diabetes (56–58). Thus, we tried to develop the ML+lab models by introducing FPG, 2hPG, or HbA1c with seemingly normal levels for diabetes screening in primary care settings. Our data indicated that introducing FPG, 2hPG, or HbA1c in individuals with normal levels of these testing increased the efficiency and accuracy of the predictive models. Moreover, the predictive accuracy of ML+FPG and ML+HbA1c models in individuals with normal levels of these testing seemed higher than that of ML model in all participants. These results suggested that it seemed practical and beneficial to adapt ML+lab models to screen out diabetes patients with seemingly normal FPG, 2hPG or HbA1c. That ML could mine maximal information

from the simple features might be an important explanation for the effectiveness of our ML+lab models screening out diabetes (59). It is reported that introducing laboratory tests, such as urine glucose, LDL-c, and triglyceride increased the predictive accuracy of diabetes predictive models by LR (45, 60), which seemed to be consistent with our findings in this study. Notably, the detection costs were decreased in the ML+FPG model and increased greatly in the ML+2hPG and ML+HbA1c models compared with the ML model, due to the testing of FPG costing much less than the testing of 2hPG and HbA1c (61). Moreover, our results indicated that the ML+FPG model had lower average detection costs compared with the NCDRS+FPG model. These results implied that the ML+FPG model had appreciable advantages in predictive accuracy and the lowest costs among the ML+lab models, the NCDRS+FPG model, and the ML model. The FPG test was most often used by health care professionals in clinical practice for it is convenient and relatively inexpensive. Therefore, the ML+FPG model was the most suitable for primary care among the ML+lab models, for its considerable predictive accuracy, low costs, and easily-accessibility, and was potentially to become a new screening strategy for diabetes in primary care with notable advantages in efficiency, economy, and convenience.

Additionally, SHAP analysis was used to perform a *post hoc* analysis of the model with all available features in the present study. Our data indicated that some predictors, including RPR, Age, SBP, WC, WHR, and BMI, were of significant value in the ML model, which is consistent with previous studies that these are important predictive factors in diabetes predictive models developed by LR and gradient boosting. Typically, the resting pulse rate is often used as an alternative to resting heart rate measurements. Increasing evidence suggested that resting heart rate was associated with type 2 diabetes (62, 63), and recent studies revealed close relevance of resting heart rate and diabetes (64, 65). Detailedly, resting heart rate is an indicator of sympathetic activation, which inhibits the insulin secretion from the pancreas and sympathetic overactivity can impair glucose uptake in skeletal muscle by inducing vasoconstriction and reducing skeletal muscle blood flow (66). These mechanisms might explain the association between RPR and diabetes. The years of cell phone use seemed relevant to age in our present study and are reported to be closely related with socioeconomic status, which might be an important explanations for the close relevance between the years of cell phone use and diabetes (67). Additionally, we identified several predictors, such as sleep duration and wake time, which are reported to be closely associated with FPG and/or HbA1c by multiple linear regression analysis, implying that more attention should be paid to these predictors in the prevention and management of diabetes (68, 69). Furthermore, our results indicated that WC, a useful measure of central obesity, was more important in predicting diabetes compared with BMI in our ML model,

which seems consistent with previous studies that WC is considered to be a more reliable measure of fat distribution and closely related with diabetes (70). It is reported that Asians including the Chinese population have more central obesity but less generalized obesity defined by high BMI. These results suggest that we should pay more attention to central obesity in the prevention, screening, and management of diabetes in the Chinese population.

Moreover, our SHAP analysis by taking two cases randomly for example identified different individualized predictors for the two selected cases, suggesting that the cutting-edge SHAP analysis in our ML model was able to screen out the crucial predictors individually for the subjects. The personalized predictors screened out would be helpful for the subject tested to get advice and take measures more accurately and precisely to prevent or treat diabetes (71).

It should be noted that we were unable to determine the detection costs and potential complication costs in the real world in the present study, although we attempted to estimate the costs using rewarding methods reported previously and adjust the costs based on economic and medical conditions in China. These findings require verification by further studies and it should be interpreted cautiously. Additionally, our data set was obtained from Han Chinese population in Hubei Province, which is located in central China, and the generalizability of our models and the findings need further testing with data from more regions and ethnic groups. Noteworthy, we had aimed to preclude the effects of medications as much as possible originally and we made great efforts to exclude the effects of antidiabetic agents on diabetes determination. Regretfully, we were not able to exclude the effects of other medications including those may affect RPR due to the missing data. In addition, the features employed in the models (e.g., RPR) were closely related with diabetes but not always play causal roles in the development of diabetes. Thus, it should be noted cautiously while interpreting the models, the importance of features, and the relationship between/among them, and further research is necessary to confirm the prevention and intervention strategies to take accordingly. Moreover, it would be of great help for us to further iterate models with new algorithms based on the newly inputted data to improve the predictive performance and generalizability of the models in the future.

Notwithstanding these limitations, the ML model developed for diabetes screening in community care had good predictive accuracy and less average detection costs compared with the NCDRS. The ML+FPG model created for diabetes screening in primary care achieved higher predictive accuracy and lower detection costs than the ML model and NCDRS+FPG model. Thus, we tentatively put forward that the ML-augmented algorithm might have the potential to become an efficient and practical tool for diabetes screening in community and primary care settings.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

LLC, XH, and XL contributed to the study design. LLC, XH, TZ, JZ, JM, ST, HZ, HH, and WZ contributed to the data acquisition and data collection. LLC, XL, XH, WZ, QZ, and PW contributed to the data analysis and the preparation of the manuscript. XL, LC, and WZ contributed to checking and modifying code related to ML algorithms. LLC and XH were responsible for the data interpretation and modifying the manuscript. All the authors have critically read the manuscript and approved the submitted version.

Funding

Acknowledgments

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fendo.2022.1043919/full#supplementary-material

References

1. Federation ID. *IDF DIABETES ATLAS (Ninth edition)*. Brussels, Belgium: International Diabetes Federation (2019).

2. Cole JB, Florez JC. Genetics of diabetes mellitus and diabetes complications. *Nat Rev Nephrol* (2020) 16(7):377–90. doi: 10.1038/s41581-020-0278-5

3. Chatterjee S, Khunti K, Davies MJ. Type 2 diabetes. *Lancet* (2017) 389 (10085):2239–51. doi: 10.1016/S0140-6736(17)30058-2

4. Kopitar L, Kocbek P, Cilar L, Sheikh A, Stiglic G. Early detection of type 2 diabetes mellitus using machine learning-based prediction models. *Sci Rep* (2020) 10(1):11981. doi: 10.1038/s41598-020-68771-z

5. Sortsø C, Komkova A, Sandbæk A, Griffin SJ, Emneus M, Lauritzen T, et al. Effect of screening for type 2 diabetes on healthcare costs: a register-based study among 139,075 individuals diagnosed with diabetes in Denmark between 2001 and 2009. *Diabetologia* (2018) 61(6):1306–14. doi: 10.1007/s00125-018-4594-2

6. Asmelash D, Asmelash Y. The burden of undiagnosed diabetes mellitus in adult African population: A systematic review and meta-analysis. *J Diabetes Res* (2019) 2019(2314-6753(2314-6753 (Electronic):4134937. doi: 10.1155/2019/4134937

7. Hu X, Zhang Q, Zeng TS, Zhang JY, Min J, Tian SH, et al. Not performing an OGTT results in underdiagnosis, inadequate risk assessment and probable cost increases of (pre)diabetes in han Chinese over 40 years: a population-based prospective cohort study. *Endocr Connect* (2018) 7(12):1507–17. doi: 10.1530/EC-18-0372

8. Zhou X, Qiao Q, Ji L, Ning F, Yang W, Weng J, et al. Nonlaboratory-based risk assessment algorithm for undiagnosed type 2 diabetes developed on a nation-wide diabetes survey. *Diabetes Care* (2013) 36(12):3944–52. doi: 10.2337/dc13-0593

9. American Diabetes A. 2. classification and diagnosis of diabetes: Standards of medical care in diabetes-2021. *Diabetes Care* (2021) 44(Suppl 1):S15–33. doi: 10.2337/dc21-S002

10. Society CD. Guideline for the prevention and treatment of type 2 diabetes mellitus in China (2020 edition). *Chin J Diabetes Mellitus* (2021) (04):315–409. doi: 10.3760/cma.j.cn115791-20210221-00095

11. Bang H, Edwards AM, Bomback AS, Ballantyne CM, Brillon D, Callahan MA, et al. Development and validation of a patient self-assessment score for diabetes risk. *Ann Intern Med* (2009) 151(11):775–83. doi: 10.7326/0003-4819-151-11-200912010-00005

12. Denisko D, Hoffman MM. Classification and interaction in random forests. *Proc Natl Acad Sci* (2018) 115(8):1690. doi: 10.1073/pnas.1800256115

13. Ngiam KY, Khor IW. Big data and machine learning algorithms for health-care delivery. *Lancet Oncol* (2019) 20(5):e262–e73. doi: 10.1016/S1470-2045(19)30149-4

14. Xu Y, Ju L, Tong J, Zhou C-M, Yang J-J. Machine learning algorithms for predicting the recurrence of stage IV colorectal cancer after tumor resection. *Sci Rep* (2020) 10(1):2519. doi: 10.1038/s41598-020-59115-y

15. Ruan Y, Bellot A, Moysova Z, Tan GD, Lumb A, Davies J, et al. Predicting the risk of inpatient hypoglycemia with machine learning using electronic health records. *Diabetes Care* (2020) 43(7):1504–11. doi: 10.2337/dc19-1743

16. Tseng PY, Chen YT, Wang CH, Chiu KM, Peng YS, Hsu SP, et al. Prediction of the development of acute kidney injury following cardiac surgery by machine learning. *Crit Care* (2020) 24(1):478. doi: 10.1186/s13054-020-03179-9

17. Vangeepuram N, Liu B, P-h C, Wang L, Pandey G. Predicting youth diabetes risk using NHANES data and machine learning. *Sci Rep* (2021) 11(1):1–9. doi: 10.1038/s41598-021-90406-0

18. Bi Y, Lu J, Wang W, Mu Y, Zhao J, Liu C, et al. Cohort profile: risk evaluation of cancers in Chinese diabetic individuals: a longitudinal (REACTION) study. *J Diabetes* (2014) 6(2):147–57. doi: 10.1111/1753-0407.12108

19. Zhang FL, Ren JX, Zhang P, Jin H, Qu Y, Yu Y, et al. Strong association of waist circumference (WC), body mass index (BMI), waist-to-Height ratio (WHtR), and waist-to-Hip ratio (WHR) with diabetes: A population-based cross-sectional study in jilin province, China. *J Diabetes Res* (2021) 2021(2314-6753(2314-6753 (Electronic):8812431. doi: 10.1155/2021/8812431

20. Bijlsma S, Bobeldijk I, Verheij ER, Ramaker R, Kochhar S, Macdonald IA, et al. Large-Scale human metabolomics studies: a strategy for data (pre-) processing and validation. *Anal Chem* (2006) 78(2):567–74. doi: 10.1021/ac051495j

21. Liang W, Liang H, Ou L, Chen B, Chen A, Li C, et al. Development and validation of a clinical risk score to predict the occurrence of critical illness in hospitalized patients with COVID-19. *JAMA Internal Med* (2020) 180(8):1081–9. doi: 10.1001/jamainternmed.2020.2033

22. Stekhoven DJ, Buhlmann P. MissForest-non-parametric missing value imputation for mixed-type data. *Bioinformatics* (2012) 28(1):112–8. doi: 10.1093/bioinformatics/btr597

23. Nguyen BP, Tay WL, Chui CK. Robust biometric recognition from palm depth images for gloved hands. *IEEE Trans Human-Machine Systems* (2015) 45(6):799–804. doi: 10.1109/THMS.2015.2453203

24. Carvalho ED, Silva RRV, Araújo FHD, Rabelo RAL, de Carvalho Filho AO. An approach to the classification of COVID-19 based on CT scans using convolutional features and genetic algorithms. *Comput Biol Med* (2021) 136:104744. doi: 10.1016/j.compbiomed.2021.104744

25. Lin D, Chen J, Lin Z, Li X, Zhang K, Wu X, et al. A practical model for the identification of congenital cataracts using machine learning. *EBioMedicine* (2020) 51:102621. doi: 10.1016/j.ebiom.2019.102621

26. Porras AR, Rosenbaum K, Tor-Diez C, Summar M, Linguraru MG. Development and evaluation of a machine learning-based point-of-care screening tool for genetic syndromes in children: a multinational retrospective study. *Lancet Digital Health* (2021) 3(10):e635–e43. doi: 10.1016/S2589-7500(21)00137-0

27. Sanderson M, Bulloch AGM, Wang J, Williams KG, Williamson T, Patten SB. Predicting death by suicide following an emergency department visit for parasuicide with administrative health care system data and machine learning. *eClinicalMedicine* (2020) 20:100281. doi: 10.1016/j.eclinm.2020.100281

28. Lundberg SM, Nair B, Vavilala MS, Horibe M, Eisses MJ, Adams T, et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nat BioMed Eng* (2018) 2(10):749–60. doi: 10.1038/s41551-018-0304-0

29. Chatterjee R, Narayan KM, Lipscomb J, Phillips LS. Screening adults for pre-diabetes and diabetes may be cost-saving. *Diabetes Care* (2010) 33(7):1484–90. doi: 10.2337/dc10-0054

30. Ye Z, Cong L, Ding G, Yu M, Zhang X, Hu R, et al. Optimal cut-off points for two-step strategy in screening of undiagnosed diabetes: A population-based study in China. *PloS One* (2014) 9(3):e87690. doi: 10.1371/journal.pone.0087690

31. Baxter M, Hudson R, Mahon J, Bartlett C, Samyshkin Y, Alexiou D, et al. Estimating the impact of better management of glycaemic control in adults with type 1 and type 2 diabetes on the number of clinical complications and the associated financial benefit. *Diabetes Med* (2016) 33(11):1575–81. doi: 10.1111/dme.13062

32. Artzi NS, Shilo S, Hadar E, Rossman H, Barbash-Hazan S, Ben-Haroush A, et al. Prediction of gestational diabetes based on nationwide electronic health records. *Nat Med* (2020) 26(1):71–6. doi: 10.1038/s41591-019-0724-8

33. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. *J Mach Learn Res* (2012) 12:2825–30. doi: 10.5555/1953048.2078195

34. Goecks J, Jalili V, Heiser LM, Gray JW. How machine learning will transform biomedicine. *Cell* (2020) 181(1):92–101. doi: 10.1016/j.cell.2020.03.022

35. Hannun AY, Rajpurkar P, Haghpanahi M, Tison GH, Bourn C, Turakhia MP, et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat Med* (2019) 25(1):65–9. doi: 10.1038/s41591-018-0268-3

36. Handelman GS, Kok HK, Chandra RV, Razavi AH, Lee MJ, Asadi H. eDoctor: machine learning and the future of medicine. *J Internal Med* (2018) 284(6):603–19. doi: 10.1111/joim.12822

37. Obermeyer Z, Emanuel EJ. Predicting the future - big data, machine learning, and clinical medicine. *N Engl J Med* (2016) 375(13):1216–9. doi: 10.1056/NEJMp1606181

38. Kumar Y, Koul A, Singla R, Ijaz MF. Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda. *J Ambient Intell Humaniz Comput* (2022) 13:1–28. doi: 10.1007/s12652-021-03612-z

39. Sharma T, Shah M. A comprehensive review of machine learning techniques on diabetes detection. *Vis Comput Ind BioMed Art.* (2021) 4(1):30. doi: 10.1186/s42492-021-00097-7

40. Letinier L, Jouganous J, Benkebil M, Bel-Letoile A, Goehrs C, Singier A, et al. Artificial intelligence for unstructured healthcare data: Application to coding of patient reporting of adverse drug reactions. *Clin Pharmacol Ther* (2021) 110(2):392–400. doi: 10.1002/cpt.2266

41. Checcucci E, Autorino R, Cacciamani GE, Amparore D, De Cillis S, Piana A, et al. Artificial intelligence and neural networks in urology: current clinical applications. *Minerva Urol Nefrol* (2020) 72(1):49–57. doi: 10.23736/S0393-2249.19.03613-0

42. Huang S, Cai N, Pacheco PP, Narrandes S, Wang Y, Xu W. Applications of support vector machine (SVM) learning in cancer genomics. *Cancer Genomics Proteomics* (2018) 15(1):41–51. doi: 10.1089/cmb.2019.0511

43. Yang L, Wu H, Jin X, Zheng P, Hu S, Xu X, et al. Study of cardiovascular disease prediction model based on random forest in eastern China. *Sci Rep* (2020) 10(1):5245. doi: 10.1038/s41598-020-62133-5

44. Moon S, Jang JY, Kim Y, Oh CM. Development and validation of a new diabetes index for the risk classification of present and new-onset diabetes: multicohort study. *Sci Rep* (2021) 11(1):15748. doi: 10.1038/s41598-021-95341-8

45. Zhang L, Wang Y, Niu M, Wang C, Wang Z. Machine learning for characterizing risk of type 2 diabetes mellitus in a rural Chinese population: the henan rural cohort study. *Sci Rep* (2020) 10(1):4406. doi: 10.1038/s41598-020-61123-x

46. Nguyen BP, Pham HN, Tran H, Nghiem N, Nguyen QH, Do TTT, et al. Predicting the onset of type 2 diabetes using wide and deep learning with electronic health records. *Comput Methods Programs Biomed* (2019) 182:105055. doi: 10.1016/j.cmpb.2019.105055

47. Wei H, Sun J, Shan W, Xiao W, Wang B, Ma X, et al. Environmental chemical exposure dynamics and machine learning-based prediction of diabetes mellitus. *Sci Total Environ* (2022) 806(Pt 2):150674. doi: 10.1016/j.scitotenv.2021.150674

48. Lai H, Huang HX, Keshavjee K, Guergachi A, Gao X. Predictive models for diabetes mellitus using machine learning techniques. *BMC Endocrine Disord* (2019) 19(1):101. doi: 10.1186/s12902-019-0436-6

49. Perveen S, Shahbaz M, Ansari MS, Keshavjee K, Guergachi A. A hybrid approach for modeling type 2 diabetes mellitus progression. *Front Genet* (2019) 10:1076. doi: 10.3389/fgene.2019.01076

50. García-Ordás MT, Benavides C, Benítez-Andrades JA, Alaiz-Moretón H, García-Rodríguez I. Diabetes detection using deep learning techniques with oversampling and feature augmentation. *Comput Methods Programs Biomed* (2021) 202:105968. doi: 10.1016/j.cmpb.2021.105968

51. Li J, Chen Q, Hu X, Yuan P, Cui L, Tu L, et al. Establishment of noninvasive diabetes risk prediction model based on tongue features and machine learning techniques. *Int J Med Informatics* (2021) 149:104429. doi: 10.1016/j.ijmedinf.2021.104429

52. Li J, Yuan P, Hu X, Huang J, Cui L, Cui J, et al. A tongue features fusion approach to predicting prediabetes and diabetes with machine learning. *J Biomed Informatics* (2021) 115:103693. doi: 10.1016/j.jbi.2021.103693

53. Deberneh HM, Kim I. Prediction of type 2 diabetes based on machine learning algorithm. *Int J Environ Res Public Health* (2021) 18(6):3317. doi: 10.3390/ijerph18063317

54. Xie Y, Nguyen QD, Hamzah H, Lim G, Bellemo V, Gunasekeran DV, et al. Artificial intelligence for teleophthalmology-based diabetic retinopathy screening in a national programme: an economic analysis modelling study. *Lancet Digit Health* (2020) 2(5):e240–e9. doi: 10.1016/S2589-7500(20)30060-1

55. American Diabetes Association Professional Practice Committee. 2. classification and diagnosis of diabetes: Standards of medical care in diabetes-2022. *Diabetes Care* (2022) 45(Suppl 1):S17–s38. doi: 10.2337/dc22-S002

56. Brambilla P, La Valle E, Falbo R, Limonta G, Signorini S, Cappellini F, et al. Normal fasting plasma glucose and risk of type 2 diabetes. *Diabetes Care* (2011) 34 (6):1372–4. doi: 10.2337/dc10-2263

57. Park SK, Ryoo JH, Oh CM, Choi JM, Choi YJ, Lee KO, et al. The risk of type 2 diabetes mellitus according to 2-h plasma glucose level: The Korean genome and epidemiology study (KoGES). *Diabetes Res Clin Pract* (2018) 146:130–7. doi: 10.1016/j.diabres.2017.08.002

58. Lu J, He J, Li M, Tang X, Hu R, Shi L, et al. Predictive value of fasting glucose, postload glucose, and hemoglobin A1c on risk of diabetes and complications in Chinese adults. *Diabetes Care* (2019) 42(8):1539–48. doi: 10.2337/dc18-1390

59. Boutilier JJ, Chan TCY, Ranjan M, Deo S. Risk stratification for early detection of diabetes and hypertension in resource-limited settings: Machine learning analysis. *J Med Internet Res* (2021) 23(1):e20123. doi: 10.2196/20123

60. Li W, Xie B, Qiu S, Huang X, Chen J, Wang X, et al. Non-lab and semi-lab algorithms for screening undiagnosed diabetes: A cross-sectional study. *EBioMedicine* (2018) 35(2352-3964(2352-3964 (Electronic):307–16. doi: 10.1016/j.ebiom.2018.08.009

61. Khunti K, Gillies CL, Taub NA, Mostafa SA, Hiles SL, Abrams KR, et al. A comparison of cost per case detected of screening strategies for type 2 diabetes and impaired glucose regulation: modelling study. *Diabetes Res Clin Pract* (2012) 97 (3):505–13. doi: 10.1016/j.diabres.2012.03.009

62. Wang L, Cui L, Wang Y, Vaidya A, Chen S, Zhang C, et al. Resting heart rate and the risk of developing impaired fasting glucose and diabetes: the kailuan prospective study. *Int J Epidemiol* (2015) 44(2):689–99. doi: 10.1093/ije/dyv079

63. Kim D-I, Yang HI, Park J-H, Lee MK, Kang D-W, Chae JS, et al. The association between resting heart rate and type 2 diabetes and hypertension in Korean adults. *Heart* (2016) 102(21):1757. doi: 10.1136/heartjnl-2015-309119

64. Yuan S, Larsson SC. An atlas on risk factors for type 2 diabetes: a wide-angled mendelian randomisation study. *Diabetologia* (2020) 63(11):2359–71. doi: 10.1007/s00125-020-05253-x

65. Guo Y, Chung W, Zhu Z, Shan Z, Li J, Liu S, et al. Genome-wide assessment for resting heart rate and shared genetics with cardiometabolic traits and type 2 diabetes. *J Am Coll Cardiol* (2019) 74(19):2162–74. doi: 10.1016/j.jacc.2019.08.1055

66. Munroe PB, Ramírez J, van Duijvenboden S. Resting heart rate and type 2 diabetes: A complex relationship in need of greater understanding. *J Am Coll Cardiol* (2019) 74(17):2175–7. doi: 10.1016/j.jacc.2019.08.1030

67. Wu H, Meng X, Wild SH, Gasevic D, Jackson CA. Socioeconomic status and prevalence of type 2 diabetes in mainland China, Hong Kong and Taiwan: a systematic review. *J Glob Health* (2017) 7(1):011103. doi: 10.7189/jogh.07.011103

68. Mokhlesi B, Temple KA, Tjaden AH, Edelstein SL, Utzschneider KM, Nadeau KJ, et al. Association of self-reported sleep and circadian measures with glycemia in adults with prediabetes or recently diagnosed untreated type 2 diabetes. *Diabetes Care* (2019) 42(7):1326–32. doi: 10.2337/dc19-0298

69. Baden MY, Hu FB, Vetter C, Schernhammer E, Redline S, Huang T. Sleep duration patterns in early to middle adulthood and subsequent risk of type 2 diabetes in women. *Diabetes Care* (2020) 43(6):1219–26. doi: 10.2337/dc19-2371

70. Lotta LA, Wittemans LBL, Zuber V, Stewart ID, Sharp SJ, Luan J, et al. Association of genetic variants related to gluteofemoral vs abdominal fat distribution with type 2 diabetes, coronary disease, and cardiovascular risk factors. *JAMA* (2018) 320(24):2553–63. doi: 10.1001/jama.2018.19329

71. Bonde A, Varadarajan KM, Bonde N, Troelsen A, Muratoglu OK, Malchau H, et al. Assessing the utility of deep neural networks in predicting postoperative surgical complications: a retrospective study. *Lancet Digit Health* (2021) 3(8):e471–e85. doi: 10.1016/S2589-7500(21)00084-4