

Published in final edited form as:

Diabetes Res Clin Pract. 2022 February 04; 185: 109237. doi:10.1016/j.diabres.2022.109237.

Population-centric Risk Prediction Modeling for Gestational Diabetes Mellitus: A Machine Learning Approach

Mukesh Kumar^{1,2,3}, Li Chen¹, Karen Tan¹, Li Ting Ang^{1,2}, Cindy Ho^{1,2}, Gerard Wong¹, Shu E Soh⁴, Kok Hian Tan^{5,6}, Jerry Kok Yen Chan^{7,8,9}, Keith M Godfrey¹⁰, Shiao-ying Chan^{1,7}, Mary Foong Fong Chong^{1,3}, John E Connolly¹¹, Yap Seng Chong^{1,7}, Johan G Eriksson^{1,7,12,13,^}, Mengling Feng^{3,*,^}, Neerja Karnani^{1,2,14,*,^}

¹Singapore Institute for Clinical Sciences, Agency for Science Technology and Research, Singapore, Republic of Singapore

²Bioinformatics Institute, Agency for Science Technology and Research, Singapore, Republic of Singapore

³Saw Swee Hock School of Public Health and Institute of Data Science, National University of Singapore and National University Health System, Singapore, Republic of Singapore

⁴Department of Paediatrics, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Republic of Singapore

⁵Division of Obstetrics and Gynecology, KK Women's and Children's Hospital, Republic of Singapore

⁶Obstetrics and Gynecology Academic Clinical Programme, Duke-NUS Graduate Medical School, Singapore, Republic of Singapore

*Co-corresponding authors: Mengling Feng: ephfm@nus.edu.sg, Neerja Karnani: neerja_karnani@sics.a-star.edu.sg.

^Joint Senior Authors: Johan G Eriksson, Mengling Feng, Neerja Karnani

Ethics approval and consent to participate

The GUSTO study have been reviewed by the National Healthcare Group (NHG) Domain Specific Review Board for ethics approval and SingHealth Centralized Institutional Review Board (CIRB/E/2019/2655). Informed consent have been obtained for the use of human biological material and/or data for future research.

Competing interests

NK, KMG, SYC and YSC are part of an academic consortium that has received research funding from Abbott Nutrition, Nestec, BenevolentAI Bio Ltd. and Danone. MF was partially supported by the National Research Foundation Singapore under its AI Singapore Programme (Award Number: [AISG-GC-2019-001-2A]). Other authors declare no conflicts of interest.

Authors' contributions

MK contributed to research study design, data curation, machine learning modeling, interpretation of results and writing of manuscript. LC contributed to GDM-IGF analysis, processing of Luminex assay data and critical reading of the manuscript. KT contributed to the acquisition, curation of adiponectin data and critical reading of the manuscript. LTA and CH contributed to clinical data curation. GW contributed to sample design, data normalization, CV correction and processing of Luminex data and critical reading of the manuscript. SES and SYC contributed to collection of phenotypic data in GUSTO cohort and critical reading of the manuscript. KHT, JKYC, KMG and YSC contributed to GUSTO cohort study design, data collection and critical reading of the manuscript. MFFC contributed to collection of dietary data in GUSTO cohort and critical reading of the manuscript. JEC contributed to Luminex assays data generation and critical reading of the manuscript. JGE contributed to interpretation of results, writing of manuscript and GUSTO cohort data collection. MF contributed to supervision of the study, interpretation of results and writing of manuscript. NK contributed to supervision of the study, interpretation of results, writing of manuscript and GUSTO cohort study data collection. MF and NK accepts full responsibility for the work, had access to the data, and controlled the decision to publish.

⁷Department of Obstetrics and Gynaecology and Human Potential Translational Research Programme, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Republic of Singapore

⁸Department of Reproductive Medicine, KK Women's and Children's Hospital, Republic of Singapore

⁹Cancer and Stem Cell Biology, Duke-NUS Medical School, Singapore, Republic of Singapore

¹⁰MRC Lifecourse Epidemiology Unit & NIHR Southampton Biomedical Research Centre, University of Southampton & University Hospital Southampton NHS Foundation Trust, UK

¹¹Institute of Molecular and Cell Biology, Agency for Science Technology and Research, Singapore, Republic of Singapore

¹²Department of General Practice and Primary Health Care, University of Helsinki, Finland

¹³Folkhälsan Research Center, Helsinki, Finland

¹⁴Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Republic of Singapore

Abstract

Aims—The heterogeneity in Gestational Diabetes Mellitus (GDM) risk factors among different populations impose challenges in developing a generic prediction model. This study evaluates the predictive ability of existing UK NICE guidelines for assessing GDM risk in Singaporean women, and used machine learning to develop a non-invasive predictive model.

Methods—Data from 909 pregnancies in Singapore's most deeply phenotyped mother-offspring cohort study, Growing Up in Singapore Towards healthy Outcomes (GUSTO), was used for predictive modeling. We used a CatBoost gradient boosting algorithm, and the Shapley feature attribution framework for model building and interpretation of GDM risk attributes.

Results—UK NICE guidelines showed poor predictability in Singaporean women [AUC:0.60 (95% CI 0.51, 0.70)]. The non-invasive predictive model comprising of 4 non-invasive factors: mean arterial blood pressure in first trimester, age, ethnicity and previous history of GDM, greatly outperformed [AUC:0.82 (95% CI 0.71, 0.93)] the UK NICE guidelines.

Conclusions—The UK NICE guidelines may be insufficient to assess GDM risk in Asian women. Our non-invasive predictive model outperforms the current state-of-the-art machine learning models to predict GDM, is easily accessible and can be an effective approach to minimize the economic burden of universal testing & GDM associated healthcare in Asian populations.

Keywords

Asian populations; Gestational Diabetes Mellitus; Heterogeneity; Machine Learning; Non-Invasive; UK NICE

Abbreviations

AI Artificial Intelligence

AUC	Area under the Receiver Operating Characteristic Curve
BMI	Body Mass Index
GCT	Glucose Challenge Test
GDM	Gestational Diabetes Mellitus
GUSTO	Growing Up in Singapore Towards healthy Outcomes
HbA_{1c}	Hemoglobin A1C
HEI-SGP	Healthy Eating Index for Pregnant women in Singapore
IADPSG	International Association of Diabetes and Pregnancy Study Groups
IDF	International Diabetes Federation
IGF1	Insulin-like Growth Factor 1
KKH	KK Women's and Children's Hospital
NICE	National Institute for Health and Care Excellence
OGTT	Oral Glucose Tolerance Test
SHAP	SHapley Additive exPlanations
WHO	World Health Organization
	Mathematical symbol delta (change in)

1 Introduction

Gestational Diabetes Mellitus (GDM) is a condition in which a woman without previous diabetes develop glucose intolerance in pregnancy [1]. This condition increases the risk of developing maternal pregnancy-induced hypertensive disorders, fetal macrosomia, caesarean birth, shoulder dystocia and other birth injuries [2]. Poorly controlled GDM also increases risks of premature birth, stillbirth and neonatal morbidity. The prevalence of GDM is increasing globally, with 1 in 6 pregnancies being affected [3]. GDM also has long-term implications for both mother and child. In the systematic reviews conducted by Vounzoulaki et al, and Kramer et al, women with GDM have been reported to have 10-fold higher risk of developing Type 2 Diabetes Mellitus, and 1.98-fold higher risk of developing cardiovascular adversities than women without GDM [4, 5]. Offspring of mothers with GDM are also at an increased risk of having metabolic adversities, perpetuating the cycle of diabetes and cardiovascular diseases [6].

Healthcare systems across the world use either the high risk selective screening approach or universal screening of GDM in pregnant women. One approach is the International Diabetes Federation (IDF) GDM Model of Care [7], designed for countries with low resources. It recommends that all pregnant women are screened at first visit by a fasting glucose, HbA_{1c} or random glucose sample, to rule out pre-existing diabetes. In those with normal early

screening, an Oral Glucose Tolerance Test (OGTT) is performed at 24-28 weeks' gestation to assess the risk of GDM. If normal, OGTT is repeated again at 32 weeks' gestation for high risk women.

The American Diabetes Association (ADA) endorses the use of either a one-step approach (IADPSG diagnostic criteria, fasting two-hour, three-point 75g OGTT) or an older two-step approach (non-fasting one-hour 50g Glucose Challenge Test (GCT), followed by diagnostic fasting three-hour 100g OGTT on a subset of women exceeding the glucose threshold value of GCT) at 24-28 weeks' gestation [8]. The two-step approach addresses the heterogeneity in populations, with varying thresholds for non-fasting one-hour 50g GCT. Either the National Diabetes Data Group (NDDG) or Carpenter and Coustan diagnostic criteria is used in fasting three-hour 100g OGTT. The population-wide benefit of one-step versus two-step approaches requires long term outcome studies.

The UK National Institute for Health and Clinical Excellence (NICE) guidelines include both research and health economic considerations for GDM. UK NICE recommends high risk selective screening for women with known GDM risk factors, such as obesity ($BMI \geq 30 \text{ kg/m}^2$), family history of diabetes, previous history of GDM, previous delivery of a macrosomic baby ($>4.5 \text{ kg}$) and ethnic origin of high diabetes prevalence (South Asian, black Caribbean or Middle Eastern) [9]. In the latest UK NICE 2015 guidelines, women with previous history of GDM are offered an OGTT at their booking appointment. Women with any of the other risk factors are offered OGTT at 24-28 weeks' gestation.

Based on findings in the Growing Up in Singapore Towards healthy Outcomes (GUSTO) study, universal screening of GDM is now recommended in Singapore [10] as the selective screening based on UK NICE 2013 guidelines failed to detect nearly half the GDM cases [11], providing evidence for heterogeneity of risk factors, hence the necessity for a population-centric approach. The International Federation of Gynecology and Obstetrics (FIGO) recommends universal GDM screening [12]. GDM screening strategies are a priority area for research, particularly in low and middle-income countries as scaling up of IDF GDM Model of Care can be resource intensive.

Machine learning algorithms have remarkable predictive power for disease stratification tasks, and hence can be beneficial in developing population based GDM risk prediction models. There are limited studies on GDM prediction using machine learning algorithms. The current state-of-the-art machine learning models are invasive and executable during pregnancy trimesters to predict GDM. Artzi et al trained a LightGBM gradient boosting classifier with Israel's Electronic Health Records (EHR) data to predict onset of GDM (AUC of 0.80 was achieved with 9 features) [13]. In another study, Wu et al trained a logistic regression classifier with China's EHR data to predict onset of GDM (AUC of 0.77 was achieved with 7 features) [14].

A non-invasive GDM risk prediction panel during early pregnancy would be the ideal counterpart to an invasive diagnostic OGTT assessment. In this study, we evaluated the predictive ability of UK NICE guidelines for GDM screening in Singapore's multi-ethnic population and developed a simple, non-invasive GDM predictive model. Our

machine learning model was implemented using prospective GUSTO cohort study data ([ClinicalTrials.gov](https://clinicaltrials.gov/ct2/show/study/NCT01174875) NCT01174875).

2 Methods

2.1 Study Population

GUSTO is a prospective multi-ethnic mother-offspring cohort study of 1,450 antenatal women recruited at 7-11 weeks of pregnancy. The pregnant mothers recruited in early pregnancy and whose children are being followed up till at least 14 years of age. There were a total of 1,344 naturally-conceived pregnancies, 96 In-Vitro Fertilization (IVF) singleton pregnancies and 10 spontaneously-conceived twin pregnancies in the cohort. Study participants were recruited from Singapore's two major public maternity hospitals; National University Hospital (NUH) and KK Women's and Children's Hospital (KKH) between June 2009 to October 2010. The participants approached were Singapore Citizens or Singapore Permanent Residents, belonging to Chinese, Malay or Indian ethnicities. Women receiving chemotherapy, psychotropic drugs or who had Type 1 Diabetes Mellitus were excluded. Only mothers who agreed to donate birth tissues at delivery (including cord, cord blood and placenta) were included.

A total of 1,166 mothers had a two-hour 75g OGTT performed (fasting, 2-hour glucose measures) in mid-gestation (median=26.9 weeks, IQR=26.4-27.6 weeks). The World Health Organization (WHO) 1999 criteria (fasting plasma glucose ≥ 7.0 mmol/L and/or 2-hour plasma glucose ≥ 7.8 mmol/L) was in use to diagnose GDM at the time of study conduct [15]. Participants of mixed ethnicity or with self-reported T2D at recruitment were excluded from model training. Analysis was restricted to 1,072 mothers whose gestation at the time of OGTT was 24¹-28⁶ weeks (gestational age is given as weeks^{days}). Supplementary Table 1 presents the statistical description of population attributes. Our models were built using 909 mothers who had complete measurements on basic physical measures, lifestyle/dietary habits, blood-derived markers and OGTT at pregnancy week 24-28 (Fig. 1).

2.2 Machine Learning

Our methodological novelty lies in combining coalitional game theory concepts with machine learning. Shapley values and the SHapley Additive exPlanations (SHAP) framework was combined with CatBoost tree ensembles for feature selection and build the population-centric GDM prediction panel [16, 17]. Lundberg and Lee have proposed SHAP as the only additive feature attribution method that satisfies two important properties of game theory - additivity (local accuracy) and monotonicity (consistency) [17].

The supervised machine learning models were built using Anaconda's distribution of Python v3.7.9 programming language in JupyterLab computational environment. The predictive models were trained using 4 machine learning algorithms to address algorithm bias; logistic regression (generalized linear model), support vector machine (linear support vector classification), CatBoost gradient boosting (tree-based) and artificial neural network (multilayer perceptron). We used fivefold stratified cross validation to preserve the same

proportion of GDM cases in each fold. Implementation details of all these models are included in the Supplementary Material [see Supplementary Material].

A machine learning model based on existing UK NICE 2015 guidelines was trained to assess the predictive ability of GDM risk in Singaporean women. The UK NICE model comprised of 5 features: pre-pregnancy obesity ($\text{BMI} \geq 30 \text{ kg/m}^2$), family history of diabetes, previous history of GDM, previous delivery of a macrosomic baby ($>=4.5 \text{ kg}$) and Indian ethnicity. We performed sensitivity analysis of UK NICE model to examine the effect of pre-pregnancy obesity thresholds in GDM prediction. We used BMI cut-offs at 23 kg/m^2 (overweight in Asian women), 25 kg/m^2 and 27.5 kg/m^2 (Obese in Asian women) [18]. We subsequently developed a two-tier predictive panel with GUSTO data. The first-tier panel was modelled using non-invasive measures at first trimester of pregnancy and the second-tier panel was modelled using additional factors gathered during mid-gestation.

2.3 Model Features

Information on demographics (maternal age, maternal ethnicity) and medical/obstetric history (self-reported pre-pregnancy weight, family history of diabetes mellitus, previous history of GDM, previous delivery of a macrosomic baby and parity) were derived from first trimester questionnaires. Systolic and diastolic blood pressure measurements were recorded at booking appointment and obtained from hospital case notes. Maternal height was measured using stadiometer, model 213; Seca, Hamburg, Germany. Around 26 weeks of pregnancy, information on self-reported smoking and alcohol consumption was collected as lifestyle habits during pregnancy.

24-hour dietary recall was retrospectively collected at 24-28 weeks of gestational age to ascertain dietary intake in a day. The nutrient analysis of dietary records was performed using Dietplan6, Forestfield Software. A relatively simple and easy to use, Healthy Eating Index for Pregnant women in Singapore (HEI-SGP) was subsequently developed to derive a diet quality score [19]. Dairy intake, total protein intake, total fat intake and total rice & alternative intake components were included for feature selection. These four individual dietary components have scores ranging from 0 to 10 points. A maximum score of 10 for dairy intake component indicates a diet rich in dairy intake.

Fasting blood plasma samples obtained at the time of the OGTT (24-28 weeks of gestational age) were analyzed for, Insulin-like Growth Factor 1 (IGF1) and adiponectin. Maternal venous blood was collected into EDTA tubes and plasma was obtained by centrifugation at $1600g$ for 10 minutes at 4° Celsius . The plasma was stored at $-80^\circ \text{ Celsius}$ until sample batch analysis. IGF1 was measured using MILLIPLEX MAP Human IGF-I, II Magnetic Bead Panel - Endocrine Assay (Merck), while adiponectin was measured using MILLIPLEX MAP Human Adipokine Magnetic Bead Panel 1 - Endocrine Multiplex Assay (Merck). Results were analysed using the Bioplex Manager 6.0 software (Biorad). We explored the hypothesis of a functional relationship between IGF1 and adiponectin protein hormones as a key endocrine modulator of metabolism [20].

3 Results

909 mothers had complete measurements on basic physical measures, lifestyle/dietary habits, blood-derived markers and OGTT at pregnancy week 24-28 (Fig. 1). The major lost to follow-up reasons after recruitment includes miscarriage and termination of pregnancy. The participant characteristics are presented in Table 1.

Fig. 2 presents the SHAP summary plot of feature selection model. GDM prediction panel was constructed using top predictors of SHAP value magnitudes more than zero. The top 9 features impacting the model outputs were mean arterial blood pressure at booking appointment, maternal age, IGF1 concentration, dairy intake, maternal height, adiponectin concentration, previous history of GDM, Malay vs Chinese/Indian ethnicity and total rice & alternatives intake.

The first-tier panel was modelled using non-invasive measures at first trimester of pregnancy: mean arterial blood pressure at booking appointment, maternal age, maternal height, previous history of GDM and Malay vs Chinese/Indian ethnicity. The second-tier panel was modelled using additional factors gathered during mid-gestation; IGF1, adiponectin, dairy intake and total rice & alternative intake. While showing data for all models in the Tables, we focus on describing the results of CatBoost machine learning models as this algorithm had the best overall performance.

The risk factors in UK NICE guidelines show poor predictive performance for GDM in Singaporean women

Table 2a presents the results of GDM prediction models trained using UK NICE 2015 guidelines. The prediction performance stagnated at AUC of 0.60 (95% CI 0.51, 0.70) in CatBoost model. In Fig. 3a, the SHAP summary plot of UK NICE CatBoost model highlights previous history of GDM as the most important GDM predictor. We additionally trained a risk stratification model based on previous history of GDM alone [AUC:0.56 (95% CI 0.53, 0.58)] (Table 3a). The addition of pre-pregnancy obesity (BMI \geq 30 kg/m²), family history of diabetes, previous delivery of a macrosomic baby and Indian ethnicity contributed to a slight boost in UK NICE model's performance.

There was a marginal improvement in CatBoost prediction performance (AUC=+0.01) when using the 23 kg/m² BMI cut-off [AUC:0.61 (95% CI 0.53, 0.69)] (Table 3a). The lowering of the obesity BMI threshold for Asian women did not improve the prediction performance of CatBoost UK NICE model. The limited predictive ability of UK NICE guidelines demonstrated by our machine learning models, substantiated the need for an improved population-centric GDM predictor.

Population-centric GDM prediction panel outperforms UK NICE guidelines

In the non-invasive panel (Table 2b), our first-tier GDM prediction panel with non-invasive features [CatBoost model AUC:0.82 (95% CI 0.71, 0.93)] outperformed the UK NICE model. The 4 features in 'NI4' model draws information from a general female population (mean arterial blood pressure at booking appointment, maternal age) to more specific

segments of the women population (with previous history of GDM, Malay vs Chinese/Indian ethnicity).

In Fig. 3b, the SHAP summary plot of 'NI4' model highlights increased mean arterial blood pressure at booking appointment as the most important feature, followed by higher maternal age, Chinese/Indian ethnicity and previous history of GDM. The SHAP plots were able to distinguish between sensitive and resilient population segments. Malay women (red feature values) had a lower risk of GDM when compared with Chinese and Indian women (blue feature values).

As shown in Table 3b, a basic 'NI2' model for the general female population (mean arterial blood pressure at booking appointment, maternal age) still outperformed the UK NICE model [CatBoost model AUC:0.75 (95% CI 0.65, 0.85)]. The inclusion of previous history of GDM in 'NI3' model increased the predictive performance [CatBoost model AUC:0.79 (95% CI 0.67, 0.91)]. The inclusion of maternal height in 'NI5' model did not improve the predictive performance of 'NI4' model.

We additionally trained the first-tier, non-invasive GDM prediction panel using a modified two-point IADPSG 2018 criteria (fasting and 2-hour glucose measures). There was a drop in the predictive performance of CatBoost 'NI4' model in two-point IADPSG 2018 criteria [AUC:0.71 (95% CI 0.62, 0.80)] (Table 3b). Despite the lack of 1-hour glucose measure in GUSTO study for full three-point IADPSG 2018 criteria, the AUC metric of 0.71 still indicates predictive power.

The addition of adiponectin at mid-gestation led to a low marginal improvement of 'NI4' model [CatBoost model AUC:0.84 (95% CI 0.75, 0.93)] (Table 2c). As seen in the 'NI4_ADI_IGFI' joint effect CatBoost model in Table 3c [AUC:0.84 (95% CI 0.76, 0.92)], addition of IGFI did not further enhance the predictive performance of 'NI4_ADI' model.

GDM prediction panel constructed at mid-gestation with adiponectin and dairy intake had the best overall performance (Table 2d) [AUC:0.85 (95% CI 0.79, 0.92)].

4 Discussion

Our findings using the UK NICE model established the need for an improved GDM predictor in an Asian population, such as in Singapore. We observed that the risk factors in UK NICE guidelines had poor GDM predictive ability for the Singapore population (AUC:0.60). The lowering of the obesity BMI thresholds applicable to Asian women did not significantly improve the UK NICE model.

We subsequently developed a two-tier GDM prediction panel that significantly outperformed the UK NICE guidelines. The first-tier GDM prediction panel is non-invasive and requires no fasting (AUC:0.82). The 4 features used in the non-invasive model can be easily measured and assessed during first trimester (mean arterial blood pressure at booking appointment, maternal age, previous history of GDM and ethnicity). Elevated mean arterial blood pressure at booking can be an early pregnancy sign of vulnerability to the metabolic syndrome of which insulin resistance and impaired glucose metabolism are prominent

components. The case-control study by Savvidou et al provides further support to our finding, where GDM women had higher systolic blood pressure in early pregnancy [21]. Hedderson et al reported similar findings, where high blood pressure in pregravid and early pregnancy states were associated with an increased risk of GDM [22]. As blood pressure is a vital sign measured routinely at antenatal visits, mean arterial blood pressure is an easy and inexpensive clinical characteristic which can be used for GDM screening.

Despite the evidence that GDM risk increases with age [23], higher maternal age is not included as one of the risk factors in UK NICE screening guidelines. This is particularly important keeping in mind that insulin resistance increases with age. With increasing age at pregnancy becoming more common in developed and developing countries, higher maternal age is an important attribute to be considered in GDM assessment.

Previous history of GDM serves as an early approach to GDM surveillance. The importance of GDM history is supported by substantial epidemiologic evidence. In a recent meta-analysis by Lee et al, women with a previous history of GDM had an 8.42-fold increased risk of developing GDM when compared with women without a previous history of GDM [24].

Studies on racial-ethnic differences in GDM risk have shown that Asians are a heterogeneous group by genetic background, culture, diet and other lifestyle factors [25]. The UK NICE guidelines classify Indian ethnic women to be at high risk for GDM in Singapore's population. In our study, we have shown that Chinese women are also at similar risk for GDM. With these findings, ethnicity-tailored preventive local programmes can be developed to improve the health literacies of GDM in high risk Chinese/Indian communities.

In our non-invasive GDM prediction panel, the addition of mean arterial blood pressure, maternal age, previous history of GDM and ethnicity resulted in a significant performance improvement ($AUC=+0.26$) when compared with the risk stratification model on previous GDM history. The 4 features in our non-invasive 'NI4' model have demonstrated stronger GDM predictive ability than the UK NICE model, suggesting that further improvements can be made in current risk assessment guidelines for GDM.

The machine learning algorithm (LightGBM gradient boosting classifier) trained by Artzi et al achieved an impressive AUC of 0.80 with 9 questionnaire features for GDM detection [13]. However, questionnaire features may introduce recall bias in predictive modelling (e.g. highest value of HbA_{1c}% measured from previous pregnancy, results of OGTT from previous pregnancy). In another study by Wu et al [14], the machine learning algorithm (logistic regression classifier) achieved an AUC of 0.77 with 7 clinical features for early GDM prediction. The invasive model developed by Wu et al requires the measurement of fasting glucose, HbA_{1c} and triglycerides. Our first-tier, non-invasive GDM prediction model has an improved performance (CatBoost model AUC:0.82) with 4 non-invasive features collected at first trimester, outperforming the current state-of-the-art machine learning models. The first-tier, non-invasive GDM prediction model can thus be an effective approach to screen and intervene early in women at risk, and also minimize the economic burden of universal testing and GDM associated healthcare in Asian populations.

The second-tier panel is invasive and requires more advanced laboratory testing, which may not be routinely available in all standard clinical laboratories. Adiponectin contributed to a better performance improvement than IGF1. With adiponectin included, the predictive performance of the non-invasive panel can only be marginally enhanced [CatBoost 'NI4_ADI' model AUC:0.84 (95% CI 0.75, 0.93)]. Lower adiponectin concentrations are associated with visceral adiposity, insulin resistance, atherosclerosis, and plays a critical role in metabolism [26]. Visceral fat accumulation is one possible pathophysiological mechanism in GDM development. Although pre-pregnancy obesity is the second most important feature in UK NICE model, pre-pregnancy obesity ($BMI \geq 30 \text{ kg/m}^2$) was of low global importance in CatBoost feature selection model (Fig. 2). As further evidenced by the stronger predictive ability of adiponectin, visceral fat accumulation (intra-abdominal fat) may be a better marker of adiposity in Asians.

Increased dairy consumption in mid-gestation added minimal predictive value to second-tier panel [CatBoost 'NI4_ADI_DI' model AUC:0.85 (95% CI 0.79, 0.92)]. Dairy consumption in GUSTO cohort study was derived from milk, yoghurt, cheese, milk-based malt drinks and cultured yoghurt drinks. Our dietary finding can be explained by general food consumption patterns during pregnancy, where dairy and dairy product consumption is greatest during mid-pregnancy. In the study by Tucker et al, high dairy intake was a strong predictor of insulin resistance in women without diabetes [27]. As mid-pregnancy is a critical window period for GDM development, dairy intake during pregnancy might be a modifiable GDM risk factor.

With the two-tier GDM prediction panel, we have shown that model prediction can be slightly enhanced by incorporating features gathered during the course of gestation. We also have a well-defined validation framework in the study as the two-tier GDM prediction panel was compared against UK NICE guidelines. An added strength of the study is the utilization of SHAP framework to interpret machine learning model outputs and design a GDM prediction panel.

This study has several limitations. Firstly, unlike large sample sizes in EHR databases, our prediction models were trained on a limited cohort of 909 pregnancies. However, EHR databases have inherent biases and are influenced by the individual's interaction with local healthcare systems. With the prospective cohort study design, GUSTO data captures the dynamic nature of complex clinical pathways and is less prone to differential measurement errors.

Secondly, the WHO 1999 GDM diagnostic criteria was in effect during two-point OGTT assessment in GUSTO study (fasting, 2-hour glucose measures). International Association of Diabetes Study Groups (IADPSG) 2018 has a less stringent criterion than WHO 1999, requiring just one abnormal glucose measure during a 2-hour 75g OGTT (fasting, 1-hour, 2-hour glucose measures). Tan et al reported that about one-third of GDM cases in KKH were diagnosed based on 1-hour glucose value [28]. The lack of 1-hour glucose measure for full three-point IADPSG 2018 criteria in GUSTO study may underestimate GDM prevalence and affect model training (AUC metric of 0.71 for modified two-point IADPSG 2018 criteria is still indicative of predictive power). As supervised machine learning models are limited by

the quality of ground truth to learn underlying patterns in data, the WHO 1999 criteria was a better ground truth labeler for training GDM algorithms using GUSTO cohort data.

Thirdly, there may be biases in the predictive value of dairy-intake in GDM risk assessment, as this measure was derived from 24-hour dietary recall. Single day intake of dietary measure is subject to recall bias and day-to-day variation. A more accurate assessment of long term dietary patterns is required in the future to build strength in the predictive value of this measure. There is also a limitation of sample size on population genomic analyses (with < 1000 samples in GUSTO study). However, key variants and iOmics analyses in GUSTO cohort have identified IGF locus and blood measures of IGF to be associated with GDM. We hence used direct measures of plasma IGF in the current analysis.

Lastly, the GUSTO cohort does not contain information of preconception parameters. With preconception data, we can possibly predict the risk of GDM during pregnancy initiation and intervene with early-stage nutritional & lifestyle changes. The longitudinal research in Singapore's PREconception study of Long-Term Maternal and Child Outcomes (S-PRESTO) birth cohort study [29], may become the basis for preconception-based GDM prediction panels to be built in the future.

Our first-tier, non-invasive predictive model would enable earlier interventions for GDM prevention and institution of earlier screening. Our machine learning tool can also be offered to pregnant women who are unwilling to have glucose challenge test taken. The trained GDM classifier can be deployed using a web application, where clinicians can enter patient information and obtain GDM risk prediction. The AI prediction model needs to be validated further using data from external cohorts or electronic health records in Singapore/Asia before deploying in local healthcare systems. A robust clinical evaluation via a randomized controlled trial is required to investigate the associations of the AI prediction tool with maternal and fetal outcomes.

5 Conclusion

Leveraging on AI, we have devised a population-based predictive care solution to assess the risk of developing GDM. The key strengths of our study lie in deep phenotyping and in applying machine learning-based predictive analytics in a prospective cohort. The state-of-the-art machine learning model can be leveraged as a rapid risk stratification tool during early pregnancy to identify Asian women at high risk of developing GDM, and implement lifestyle interventions. The translational impact of this unique Asian study would transform women's health: shifting from a reactive to predictive care strategy in GDM management.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank the GUSTO study team for their help in acquiring the research data and their crucial work with the participants.

Funding

The GUSTO birth cohort study is supported by the Translational Clinical Research (TCR) Flagship Program on Developmental Pathways to Metabolic Disease and Open Fund Large Collaborative Grant (OFLCG) Programmes, funded by the National Research Foundation (NRF) and administered by the National Medical Research Council (NMRC), Singapore (award numbers NMRC/TCR/004-NUS/2008, NMRC/TCR/012-NUHS/2014, OFLCG/MOH-000504). This research is supported by NMRC's Open Fund - Large Collaborative Grant, titled 'Metabolic Health in Asian Women and their Children' (award number OFLCG19may-0033). KMG is supported by the UK Medical Research Council (MC_UU_12011/4), the National Institute for Health Research (NIHR Senior Investigator (NF-SI-0515-10042) and NIHR Southampton Biomedical Research Centre (IS-BRC-1215-20004)) and the British Heart Foundation (RG/15/17/3174). Additional funds for data analysis were supported by the Strategic Positioning Fund and IAFpp funds (H17/01/a0/005) available to NK through Agency for Science, Technology and Research (A*STAR), Singapore (award number SPF 002/2013).

Availability of data and materials

The data that support the findings of this research are available from the corresponding authors upon reasonable request.

The code generated to reproduce this research is available at Github page: https://github.com/mukeshkumar/GUSTO_Gestational-Diabetes-Mellitus

The non-invasive GDM predictive model (CatBoost algorithm) have been deployed into a web application and can be accessed through the following URL: <https://www.morninfeng.com/all-projects-and-demos#gdm>

References

- [1]. Metzger BE, Coustan DR. Summary and recommendations of the Fourth International Workshop-Conference on Gestational Diabetes Mellitus. The Organizing Committee. *Diabetes Care*. 1998; 21 (Suppl 2) B161–7. [PubMed: 9704245]
- [2]. Gestational Diabetes Mellitus. *Diabetes Care*. 2003; 26: s103–s5. [PubMed: 12502631]
- [3]. International Diabetes Federation. *IDF Diabetes Atlas*. 9th edn. 2019.
- [4]. Vounzoulaki E, Khunti K, Abner SC, Tan BK, Davies MJ, Gillies CL. Progression to type 2 diabetes in women with a known history of gestational diabetes: systematic review and meta-analysis. *BMJ*. 2020; 369 m1361 [PubMed: 32404325]
- [5]. Kramer CK, Campbell S, Retnakaran R. Gestational diabetes and the risk of cardiovascular disease in women: a systematic review and meta-analysis. *Diabetologia*. 2019; 62: 905–14. [PubMed: 30843102]
- [6]. Chu AHY, Godfrey KM. Gestational Diabetes Mellitus and Developmental Programming. *Annals of Nutrition and Metabolism*. 2020; 76 (suppl 3) 4–15.
- [7]. International Diabetes Federation. *IDF GDM Model of Care*. 2015.
- [8]. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes - 2020. *Diabetes Care*. 2020; 43: S14–S31. [PubMed: 31862745]
- [9]. National Institute for Health and Care Excellence. *Diabetes in pregnancy: management from preconception to the postnatal period*. 2015.
- [10]. Academy of Medicine Singapore. *Guidelines for the Management of Gestational Diabetes Mellitus*. 2018.
- [11]. Chong YS, Cai S, Lin H, Soh SE, Lee YS, Leow MK, et al. Ethnic differences translate to inadequacy of high-risk screening for gestational diabetes mellitus in an Asian population: a cohort study. *BMC Pregnancy Childbirth*. 2014; 14: 345. [PubMed: 25273851]
- [12]. Hod M, Kapur A, Sacks DA, Hadar E, Agarwal M, Di Renzo GC, et al. The International Federation of Gynecology and Obstetrics (FIGO) Initiative on gestational diabetes mellitus: A pragmatic guide for diagnosis, management, and care. *Int J Gynaecol Obstet*. 2015; 131 (Suppl 3) S173–211.

- [13]. 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: 71–6. [PubMed: 31932807]
- [14]. Wu Y-T, Zhang C-J, Mol BW, Kawai A, Li C, Chen L, et al. Early Prediction of Gestational Diabetes Mellitus in the Chinese Population via Advanced Machine Learning. *The Journal of Clinical Endocrinology & Metabolism*. 2020; 106: e1191–e205.
- [15]. World Health Organization. *Diagnostic Criteria and Classification of Hyperglycaemia First Detected in Pregnancy*. 2013.
- [16]. Prokhorenkova L, Gusev G, Vorobev A, Veronika Dorogush A, Gulin A. CatBoost: unbiased boosting with categorical features. *arXiv e-prints*. 2017.
- [17]. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*. 2020; 2: 56–67.
- [18]. Consultation WHOE. Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. *Lancet*. 2004; 363: 157–63. [PubMed: 14726171]
- [19]. Han CY, Colega M, Quah EPL, Chan YH, Godfrey KM, Kwek K, et al. A healthy eating index to measure diet quality in pregnant women in Singapore: a cross-sectional study. *BMC Nutrition*. 2015; 1: 39.
- [20]. Orrù S, Nigro E, Mandola A, Alfieri A, Buono P, Daniele A, et al. A Functional Interplay between IGF-1 and Adiponectin. *Int J Mol Sci*. 2017; 18 2145
- [21]. Savvidou M, Nelson SM, Makgoba M, Messow CM, Sattar N, Nicolaides K. First-trimester prediction of gestational diabetes mellitus: examining the potential of combining maternal characteristics and laboratory measures. *Diabetes*. 2010; 59: 3017–22. [PubMed: 20876721]
- [22]. Hedderon MM, Ferrara A. High Blood Pressure Before and During Early Pregnancy Is Associated With an Increased Risk of Gestational Diabetes Mellitus. *Diabetes Care*. 2008; 31: 2362–7. [PubMed: 18809624]
- [23]. Khalil A, Syngelaki A, Maiz N, Zinevich Y, Nicolaides KH. Maternal age and adverse pregnancy outcome: a cohort study. *Ultrasound in Obstetrics & Gynecology*. 2013; 42: 634–43. [PubMed: 23630102]
- [24]. Lee KW, Ching SM, Ramachandran V, Yee A, Hoo FK, Chia YC, et al. Prevalence and risk factors of gestational diabetes mellitus in Asia: a systematic review and meta-analysis. *BMC Pregnancy Childbirth*. 2018; 18: 494. [PubMed: 30547769]
- [25]. Chu SY, Abe K, Hall LR, Kim SY, Njoroge T, Qin C. Gestational diabetes mellitus: All Asians are not alike. *Preventive Medicine*. 2009; 49: 265–8. [PubMed: 19596364]
- [26]. Swarbrick MM, Havel PJ. Physiological, pharmacological, and nutritional regulation of circulating adiponectin concentrations in humans. *Metab Syndr Relat Disord*. 2008; 6: 87–102. [PubMed: 18510434]
- [27]. Tucker LA, Erickson A, LeCheminant JD, Bailey BW. Dairy consumption and insulin resistance: the role of body fat, physical activity, and energy intake. *J Diabetes Res*. 2015; 2015 206959 [PubMed: 25710041]
- [28]. Tan, KH; Kwek, K; Ng, MJ; Yeo, SH; Wright, A; Tagore, S; , et al. Introduction of the IADPSG Criteria for Screening and Diagnosis of Gestational Diabetes Mellitus; SingHealth Duke-NUS Scientific Congress 2016; 2016.
- [29]. Loo E, Soh S-E, Loy S, Ng S, Tint M, Chan S-Y, et al. Cohort profile: Singapore Preconception Study of Long-Term Maternal and Child Outcomes (S-PRESTO). *European Journal of Epidemiology*. 2020; 1

- UK NICE guidelines may be insufficient to assess GDM risk in Asian women.
- Higher blood pressure during first trimester is a top risk factor for GDM.
- Non-invasive AI model can be leveraged as a rapid GDM risk stratification tool.
- AI model is robust when using a modified two-point IADPSG 2018 GDM criteria.
- AI model may be a cost effective alternative strategy to universal GDM screening.

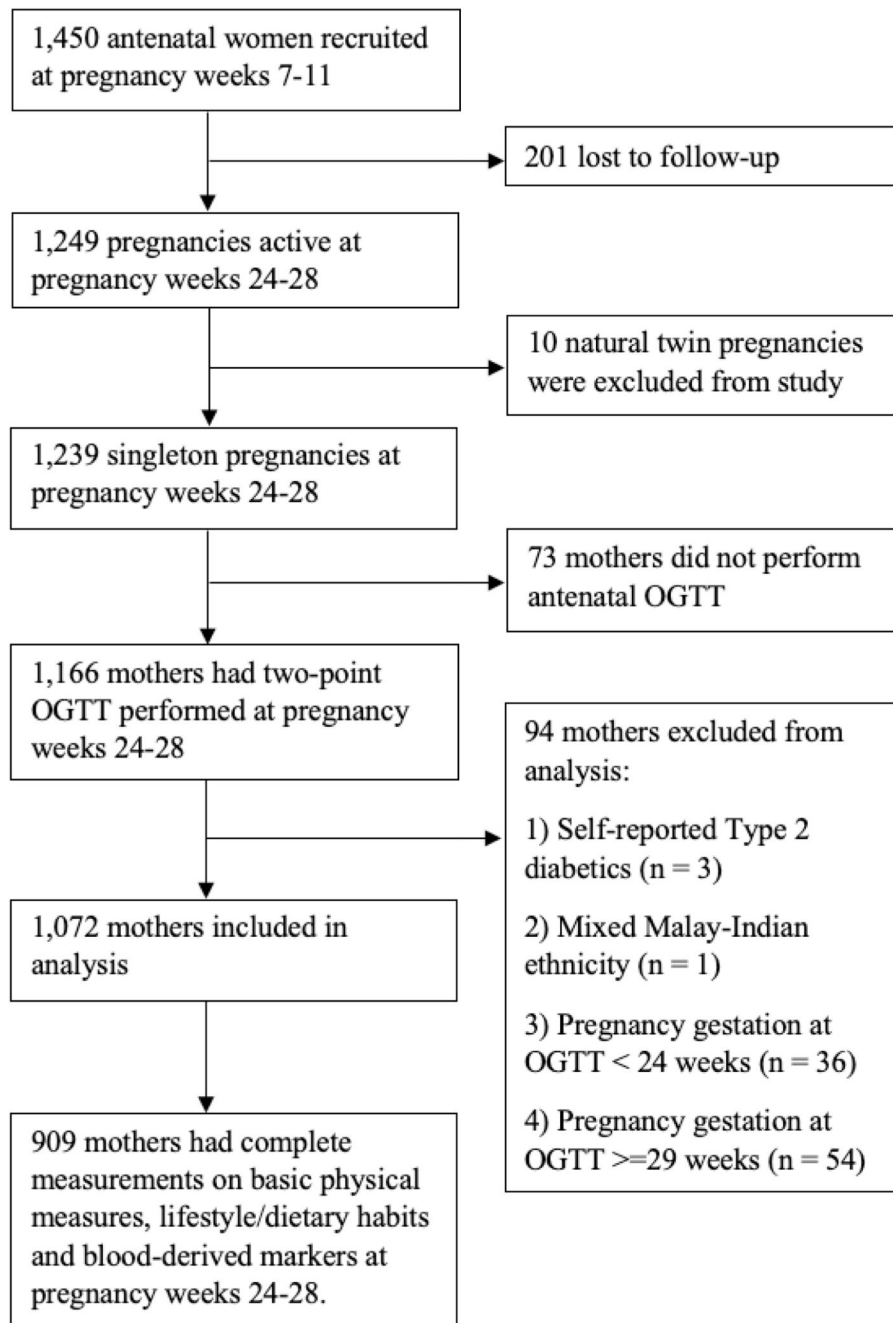


Fig. 1. Population selection flowchart of 909 mothers who had complete measurements on basic physical measures, lifestyle/dietary habits, blood-derived markers and OGTT at pregnancy week 24-28 for machine learning models.

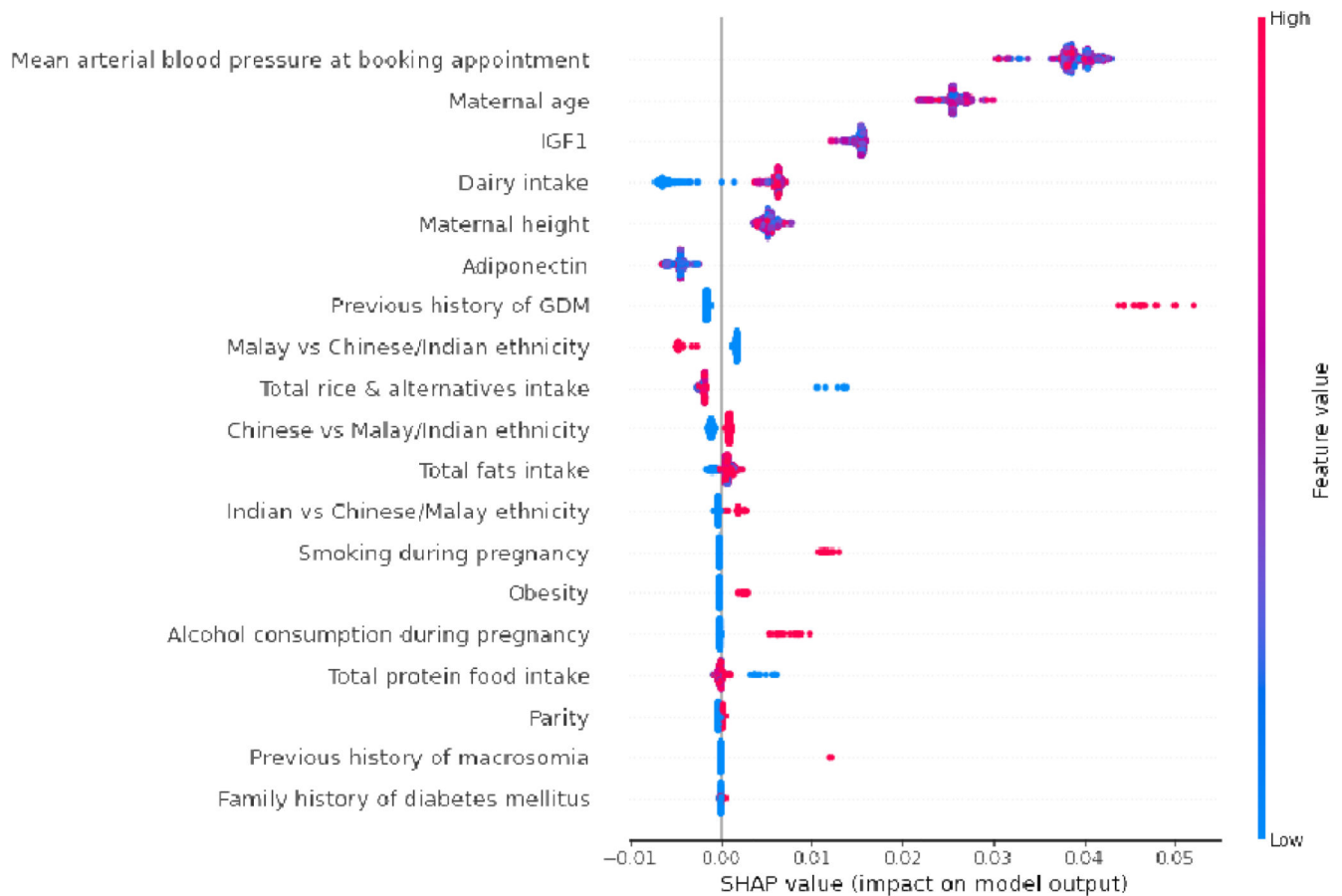
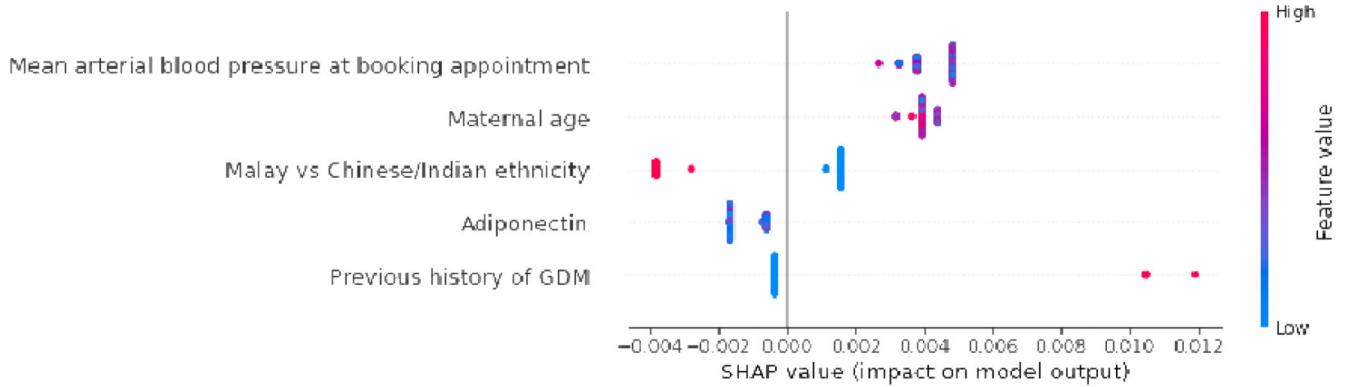
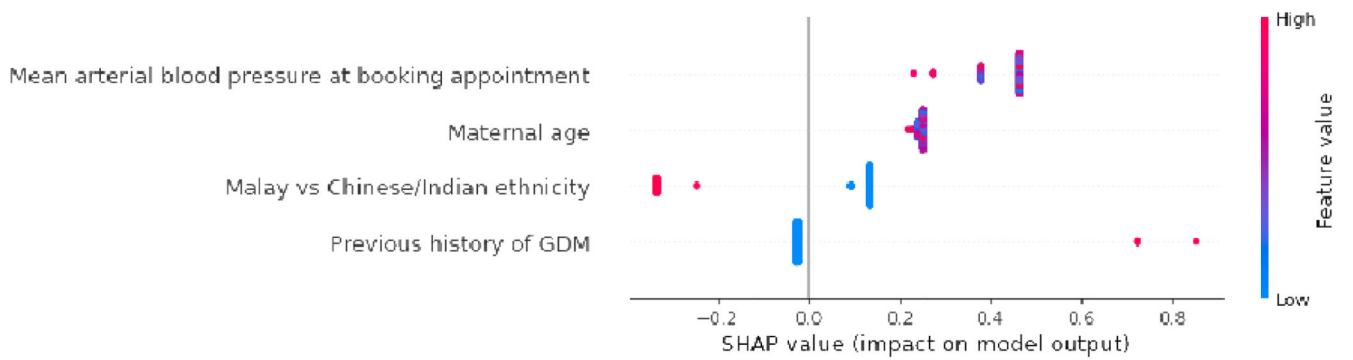
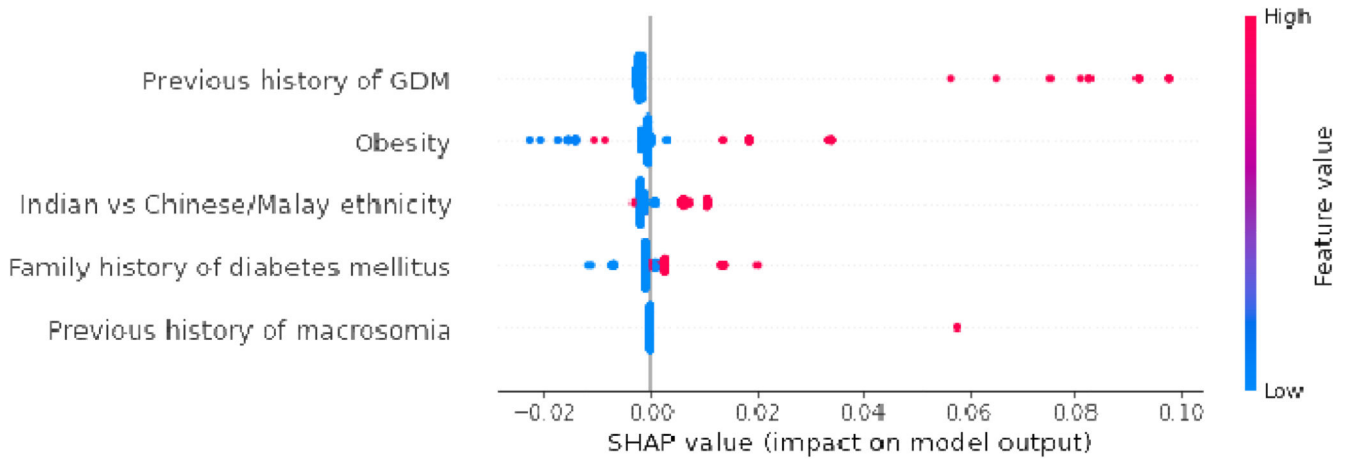


Fig. 2. Global importance of individual features and their correlation with GDM/Non-GDM outcomes estimated using the Shapley values computed from coalitional game theory. Local explanations are plotted in a beeswarm style to observe the magnitude and prevalence of a feature's effects, where each subject have one dot for each feature. High positive SHAP values drive model predictions for GDM and low negative SHAP values drive model predictions for non-GDM, as predicted by CatBoost feature selection model. SHAP values close to zero means that the feature contributes little to the prediction. The colored feature values represents the range of values taken by individual features. Blue corresponds to low feature value and red corresponds to high feature value (direction of feature effects). For example, higher mean arterial blood pressure values lead to positive SHAP values (red feature values) and drive model predictions for GDM outcome, whereas lower dairy intake values lead to negative SHAP values (blue feature values) and drive model predictions for non-GDM outcome. Similarly, in case of binary features such as previous history of GDM (0: No, 1: Yes) and Malay vs Chinese/Indian ethnicity (0: Chinese or Indian, 1: Malay), the blue features represents feature value 0 and red corresponds to feature value 1. SHAP values represent a change in log odds ratio and features are sorted by global impact. If the feature's impact (value changes) varies smoothly on the model's output, the coloring will have a smooth gradation like mean arterial blood pressure. As observed in maternal height,

multiple dots at the same position in the horizontal axis are piled up and shown as density.
(Color should be used for figure in print)



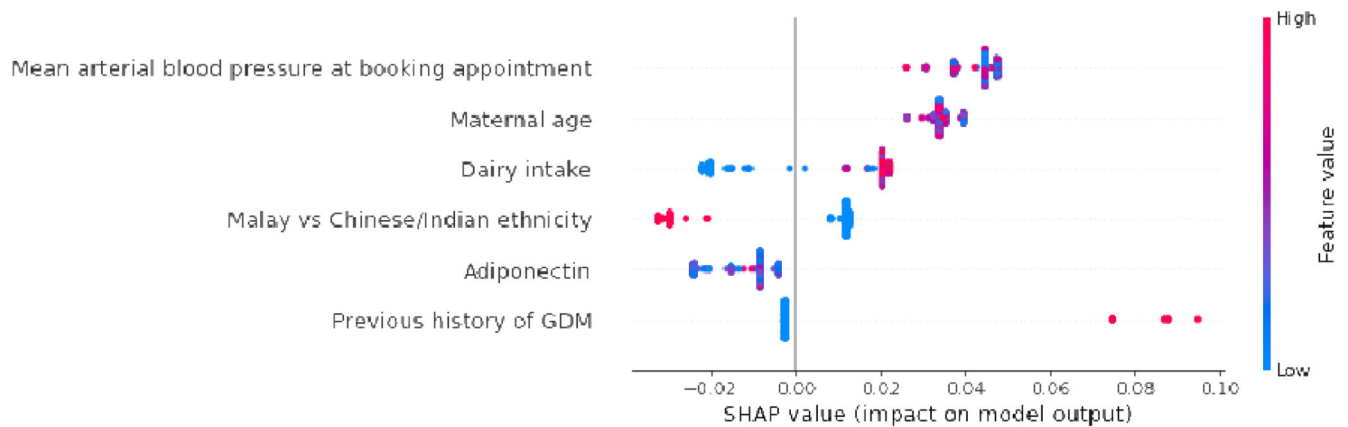


Fig. 3.

a: Global importance of individual features and their correlation with GDM/Non-GDM outcomes estimated using the Shapley values computed from coalitional game theory. Local explanations are plotted in a beeswarm style to observe the magnitude and prevalence of a feature's effects, where each subject have one dot for each feature. High positive SHAP values drive model predictions for GDM and low negative SHAP values drive model predictions for non-GDM, as predicted by CatBoost UK NICE model. SHAP values close to zero means that the feature contributes little to the prediction. The colored feature values represents the range of values taken by individual features. Blue corresponds to low feature value and red corresponds to high feature value (direction of feature effects). In binary features such as previous history of GDM (0: No, 1: Yes) and Indian vs Chinese/Malay ethnicity (0: Chinese or Malay, 1: Indian), the blue features represents feature value 0 and red corresponds to feature value 1. Indian women (red feature values) had a higher risk of GDM when compared with Chinese and Malay women (blue feature values). SHAP values represent a change in log odds ratio and features are sorted by global impact. If the feature's impact (value changes) varies smoothly on the model's output, the coloring will have a smooth gradation like subjects with previous history of GDM (red feature values). (Color should be used for figure in print)

b: Global importance of individual features and their correlation with GDM/Non-GDM outcomes estimated using the Shapley values computed from coalitional game theory. Local explanations are plotted in a beeswarm style to observe the magnitude and prevalence of a feature's effects, where each subject have one dot for each feature. High positive SHAP values drive model predictions for GDM and low negative SHAP values drive model predictions for non-GDM, as predicted by CatBoost NI4 model. SHAP values close to zero means that the feature contributes little to the prediction. The colored feature values represents the range of values taken by individual features. Blue corresponds to low feature value and red corresponds to high feature value (direction of feature effects). For example, higher mean arterial blood pressure values lead to positive SHAP values (red feature values) and drive model predictions for GDM outcome. Similarly, in binary features such as previous history of GDM (0: No, 1: Yes) and Malay vs Chinese/Indian ethnicity (0: Chinese or Indian, 1: Malay), the blue features represents feature value 0 and red corresponds to feature value 1. Malay women (red feature values) had a lower risk of GDM when compared with Chinese and Indian women (blue feature values). SHAP values represent a change

in log odds ratio and features are sorted by global impact. If the feature's impact (value changes) varies smoothly on the model's output, the coloring will have a smooth gradation like mean arterial blood pressure. (Color should be used for figure in print)

c: Global importance of individual features and their correlation with GDM/Non-GDM outcomes estimated using the Shapley values computed from coalitional game theory. Local explanations are plotted in a beeswarm style to observe the magnitude and prevalence of a feature's effects, where each subject have one dot for each feature. High positive SHAP values drive model predictions for GDM and low negative SHAP values drive model predictions for non-GDM, as predicted by CatBoost NI4_ADI model. SHAP values close to zero means that the feature contributes little to the prediction. The colored feature values represents the range of values taken by individual features. Blue corresponds to low feature value and red corresponds to high feature value (direction of feature effects). For example, higher mean arterial blood pressure values lead to positive SHAP values (red feature values) and drive model predictions for GDM outcome. Similarly, in binary features such as previous history of GDM (0: No, 1: Yes) and Malay vs Chinese/Indian ethnicity (0: Chinese or Indian, 1: Malay), the blue features represents feature value 0 and red corresponds to feature value 1. Malay women (red feature values) had a lower risk of GDM when compared with Chinese and Indian women (blue feature values). SHAP values represent a change in log odds ratio and features are sorted by global impact. If the feature's impact (value changes) varies smoothly on the model's output, the coloring will have a smooth gradation like mean arterial blood pressure. (Color should be used for figure in print)

d: Global importance of individual features and their correlation with GDM/Non-GDM outcomes estimated using the Shapley values computed from coalitional game theory. Local explanations are plotted in a beeswarm style to observe the magnitude and prevalence of a feature's effects, where each subject have one dot for each feature. High positive SHAP values drive model predictions for GDM and low negative SHAP values drive model predictions for non-GDM, as predicted by CatBoost NI4_ADI_DI model. SHAP values close to zero means that the feature contributes little to the prediction. The colored feature values represents the range of values taken by individual features. Blue corresponds to low feature value and red corresponds to high feature value (direction of feature effects). For example, higher mean arterial blood pressure values lead to positive SHAP values (red feature values) and drive model predictions for GDM outcome, whereas lower dairy intake values lead to negative SHAP values (blue feature values) and drive model predictions for non-GDM outcome. Similarly, in binary features such as previous history of GDM (0: No, 1: Yes) and Malay vs Chinese/Indian ethnicity (0: Chinese or Indian, 1: Malay), the blue features represents feature value 0 and red corresponds to feature value 1. Malay women (red feature values) had a lower risk of GDM when compared with Chinese and Indian women (blue feature values). SHAP values represent a change in log odds ratio and features are sorted by global impact. If the feature's impact (value changes) varies smoothly on the model's output, the coloring will have a smooth gradation like mean arterial blood pressure. (Color should be used for figure in print)

Table 1

Participant Characteristics

	GUSTO (n = 909)
<u>Demographics</u>	
Age (years), mean \pm SD	30.3 \pm 5.1
Ethnicity, n (%)	
Chinese	507 (55.8)
Malay	241 (26.5)
Indian	161 (17.7)
<u>Pregnancy/Medical History</u>	
Pre-pregnancy BMI (kg/m ²), mean \pm SD	22.7 \pm 4.5
Family history of diabetes mellitus, n (%)	
Yes	278 (30.6)
No	631 (69.4)
Previous history of GDM, n (%)	
Yes	24 (2.6)
No	885 (97.4)
Previous history of macrosomia, n (%)	
Yes	20(0.2)
No	907 (99.8)
Parity, n (%)	
Nulliparous	404 (44.4)
Multiparous	505 (55.6)
<u>Basic Physical Measures</u>	
Mean arterial blood pressure at booking appointment (mmHg), mean \pm SD	82.7 \pm 9.6
Maternal height (cm), mean \pm SD	158.2 \pm 5.7
<u>Lifestyle/Dietary Habits</u>	
Self-reported smoking during pregnancy, n (%)	
Yes	25 (2.8)
No	884 (97.2)
Self-reported alcohol consumption during pregnancy, n (%)	

	GUSTO (n = 909)
Yes	23 (2.5)
No	886 (97.5)
Healthy Eating Index of diary intake from 24-hr recall (10 as the maximum score), mean \pm SD	5.4 \pm 4.1
Healthy Eating Index of total protein foods intake from 24-hr recall (10 as the maximum score), mean \pm SD	8.1 \pm 2.8
Healthy Eating Index of total fats intake from 24-hr recall (10 as the maximum score), mean \pm SD	6.1 \pm 3.9
Healthy Eating Index of total rice & alternatives intake from 24-hr recall (10 as the maximum score), mean \pm SD	8.7 \pm 1.9
<u>Blood-Derived Markers</u>	
IGF1 (pg/ml), mean \pm SD	5.53 x 10 ⁴ \pm 2.39 x 10 ⁴
Adiponectin (pg/ml), mean \pm SD	1.76 x 10 ⁷ \pm 1.30 x 10 ⁷
<u>OGTT at 24¹-28⁶ weeks' gestation</u>	
Glucose measures (mmol/L), mean \pm SD	
Fasting glucose	4.3 \pm 0.4
2-hour glucose	6.5 \pm 1.4
GDM, n (%)	
WHO 1999 criteria	154 (16.9)
Two-point IADPSG 2018 criteria	106 (11.7)

Table 2a

Baseline GDM Prediction Models (UK NICE)

Features (UK NICE)	Pre-pregnancy obesity (30 kg/m ²) + Family history of diabetes + Previous history of GDM + Previous delivery of macrosomic baby + Indian ethnicity
Model	AUC (95% CI)
Logistic Regression	0.59 (0.52, 0.65)
Support Vector Machine	0.60 (0.50, 0.70)
Neural Network	0.61 (0.50, 0.72)
CatBoost	0.60 (0.51, 0.70)

Table 2b

First-Tier GDM Prediction Panel (Non-Invasive)

Features (N14)	Mean arterial blood pressure at booking visit + Maternal age + Previous history of GDM + Ethnicity
Model	AUC (95% CI)
Logistic Regression	0.83 (0.74, 0.92)
Support Vector Machine	0.83 (0.74, 0.92)
Neural Network	0.83 (0.75, 0.92)
CatBoost	0.82 (0.71, 0.93)

Table 2c
Second-Tier GDM Prediction Panel (Non-Invasive & Blood-Derived Marker)

Features (NI4_ADD)	Mean arterial blood pressure at booking visit + Maternal age + Previous history of GDM + Ethnicity + Adiponectin
Model	AUC (95% CI)
Logistic Regression	0.83 (0.75, 0.91)
Support Vector Machine	0.83 (0.75, 0.90)
Neural Network	0.84 (0.77, 0.90)
CatBoost	0.84 (0.75, 0.93)

Table 2d
Second-Tier GDM Prediction Panel (Non-Invasive & Blood-Derived Marker with Maternal Diet)

Features (NI4_ADI_DI)	Mean arterial blood pressure at booking visit + Maternal age + Previous history of GDM + Ethnicity + Adiponectin + Dairy Intake
Model	AUC (95% CI)
Logistic Regression	0.85 (0.81, 0.89)
Support Vector Machine	0.84 (0.80, 0.89)
Neural Network	0.85 (0.80, 0.89)
CatBoost	0.85 (0.79, 0.92)

Table 3a

Baseline GDM Prediction Models (UK NICE)

Features (PHGDM)	Previous history of GDM
Model	AUC (95% CI)
Logistic Regression	0.56 (0.53, 0.58)
Support Vector Machine	0.56 (0.53, 0.58)
Neural Network	0.56 (0.53, 0.58)
CatBoost	0.56 (0.53, 0.58)
Features (UK NICE: 27.5 kg/m² BMI cut-off)	Pre-pregnancy obesity (27.5 kg/m ²) + Family history of diabetes + Previous history of GDM + Previous delivery of macrosomic baby + Indian ethnicity
Model	AUC (95% CI)
Logistic Regression	0.57 (0.48, 0.66)
Support Vector Machine	0.62 (0.47, 0.76)
Neural Network	0.60 (0.52, 0.68)
CatBoost	0.56 (0.42, 0.70)
Features (UK NICE: 25 kg/m² BMI cut-off)	Pre-pregnancy obesity (25 kg/m ²) + Family history of diabetes + Previous history of GDM + Previous delivery of macrosomic baby + Indian ethnicity
Model	AUC (95% CI)
Logistic Regression	0.55 (0.47, 0.62)
Support Vector Machine	0.61 (0.42, 0.79)
Neural Network	0.58 (0.48, 0.68)
CatBoost	0.55 (0.44, 0.66)
Features (UK NICE: 23 kg/m² BMI cut-off)	Pre-pregnancy obesity (23 kg/m ²) + Family history of diabetes + Previous history of GDM + Previous delivery of macrosomic baby + Indian ethnicity
Model	AUC (95% CI)
Logistic Regression	0.59 (0.50, 0.68)
Support Vector Machine	0.59 (0.31, 0.87)
Neural Network	0.62 (0.46, 0.78)
CatBoost	0.61 (0.53, 0.69)

Table 3b

First-Tier GDM Prediction Panel (Non-Invasive)

Features (NI2)	Mean arterial blood pressure at booking visit + Maternal age
Model	AUC (95% CI)
Logistic Regression	0.76 (0.68, 0.84)
Support Vector Machine	0.76 (0.69, 0.83)
Neural Network	0.76 (0.66, 0.86)
CatBoost	0.75 (0.65, 0.85)
Features (NI3)	Mean arterial blood pressure at booking visit + Maternal age + Previous history of GDM
Model	AUC (95% CI)
Logistic Regression	0.79 (0.70, 0.87)
Support Vector Machine	0.79 (0.70, 0.88)
Neural Network	0.79 (0.70, 0.88)
CatBoost	0.79 (0.67, 0.91)
Features (NI5)	Mean arterial blood pressure at booking visit + Maternal age + Previous history of GDM + Ethnicity + Maternal height
Model	AUC (95% CI)
Logistic Regression	0.83 (0.75, 0.91)
Support Vector Machine	0.83 (0.76, 0.89)
Neural Network	0.83 (0.77, 0.89)
CatBoost	0.82 (0.72, 0.92)
Features (NI4: modified two-point IADPSG 2018 criteria)	Mean arterial blood pressure at booking visit + Maternal age + Previous history of GDM + Ethnicity
Model	AUC (95% CI)
Logistic Regression	0.72 (0.62, 0.81)
Support Vector Machine	0.71 (0.62, 0.80)
Neural Network	0.73 (0.68, 0.79)
CatBoost	0.71 (0.62, 0.80)

Table 3c
Second-Tier GDM Prediction Panel (Non-Invasive & Blood-Derived Marker, Non-Invasive & Maternal Diet)

Features (NI4_IGFI)	Mean arterial blood pressure at booking visit + Maternal age + Previous history of GDM + Ethnicity + IGFI
Model	AUC (95% CI)
Logistic Regression	0.83 (0.76, 0.90)
Support Vector Machine	0.82 (0.75, 0.90)
Neural Network	0.83 (0.76, 0.89)
CatBoost	0.83 (0.73, 0.93)
Features (NI4_ADI_IGFI)	Mean arterial blood pressure at booking visit + Maternal age + Previous history of GDM + Ethnicity + Adiponectin + IGFI
Model	AUC (95% CI)
Logistic Regression	0.83 (0.76, 0.90)
Support Vector Machine	0.82 (0.76, 0.89)
Neural Network	0.83 (0.74, 0.93)
CatBoost	0.84 (0.76, 0.92)
Features (NI4_DI)	Mean arterial blood pressure at booking visit + Maternal age + Previous history of GDM + Ethnicity + Dairy Intake
Model	AUC (95% CI)
Logistic Regression	0.85 (0.81, 0.89)
Support Vector Machine	0.84 (0.78, 0.90)
Neural Network	0.85 (0.79, 0.90)
CatBoost	0.84 (0.77, 0.92)
Features (NI4_CI)	Mean arterial blood pressure at booking visit + Maternal age + Previous history of GDM + Ethnicity + Total Rice & Alternatives Intake
Model	AUC (95% CI)
Logistic Regression	0.82 (0.74, 0.90)
Support Vector Machine	0.82 (0.73, 0.90)
Neural Network	0.82 (0.74, 0.90)
CatBoost	0.82 (0.73, 0.90)
Features (NI4_DCI)	Mean arterial blood pressure at booking visit + Maternal age + Previous history of GDM + Ethnicity + Dairy Intake + Total Rice & Alternatives Intake
Model	AUC (95% CI)
Logistic Regression	0.84 (0.77, 0.90)
Support Vector Machine	0.83 (0.76, 0.90)

Features (NI4_IGFI)	Mean arterial blood pressure at booking visit + Maternal age + Previous history of GDM + Ethnicity + IGFI
Neural Network	0.82 (0.74, 0.90)
CatBoost	0.84 (0.76, 0.91)

Table 3d
Second-Tier GDM Prediction Panel (Non-Invasive & Blood-Derived Marker with Maternal Diet)

Features (N14_ADI_CI)	Mean arterial blood pressure at booking visit + Maternal age + Previous history of GDM + Ethnicity + Adiponectin + Total Rice & Alternatives Intake
Model	AUC (95% CI)
Logistic Regression	0.82 (0.74, 0.90)
Support Vector Machine	0.82 (0.74, 0.90)
Neural Network	0.82 (0.74, 0.90)
CatBoost	0.83 (0.76, 0.91)
Features (N14_ADI_DCI)	Mean arterial blood pressure at booking visit + Maternal age + Previous history of GDM + Ethnicity + Adiponectin + Dairy Intake + Total Rice & Alternatives Intake
Model	AUC (95% CI)
Logistic Regression	0.84 (0.79, 0.89)
Support Vector Machine	0.84 (0.77, 0.90)
Neural Network	0.82 (0.75, 0.90)
CatBoost	0.85 (0.78, 0.92)