scientific reports

OPEN

MIDO GDM: an innovative artificial intelligence-based prediction model for the development of gestational diabetes in Mexican women

Héctor Gallardo-Rincón[1,2], María Jesús Ríos-Blancas[2,3], Janinne Ortega-Montiel[2], Alejandra Montoya[2], Luis Alberto Martinez-Juarez[2✉], Julieta Lomelín-Gascón[2], Rodrigo Saucedo-Martínez[2], Ricardo Mújica-Rosales[2], Victoria Galicia-Hernández[2], Linda Morales-Juárez[2], Lucía Marcela Illescas-Correa[4], Ixel Lorena Ruiz-Cabrera[4], Daniel Alberto Díaz-Martínez[5], Francisco Javier Magos-Vázquez[5], Edwin Oswaldo Vargas Ávila[5], Alejandro Efraín Benitez-Herrera[6], Diana Reyes-Gómez[6], María Concepción Carmona-Ramos[6], Laura Hernández-González[6], Oscar Romero-Islas[6], Enrique Reyes Muñoz[7] & Roberto Tapia-Conyer[8]

Given the barriers to early detection of gestational diabetes mellitus (GDM), this study aimed to develop an artificial intelligence (AI)-based prediction model for GDM in pregnant Mexican women. Data were retrieved from 1709 pregnant women who participated in the multicenter prospective cohort study 'Cuido mi embarazo'. A machine-learning-driven method was used to select the best predictive variables for GDM risk: age, family history of type 2 diabetes, previous diagnosis of hypertension, pregestational body mass index, gestational week, parity, birth weight of last child, and random capillary glucose. An artificial neural network approach was then used to build the model, which achieved a high level of accuracy (70.3%) and sensitivity (83.3%) for identifying women at high risk of developing GDM. This AI-based model will be applied throughout Mexico to improve the timing and quality of GDM interventions. Given the ease of obtaining the model variables, this model is expected to be clinically strategic, allowing prioritization of preventative treatment and promising a paradigm shift in prevention and primary healthcare during pregnancy. This AI model uses variables that are easily collected to identify pregnant women at risk of developing GDM with a high level of accuracy and precision.

Gestational diabetes mellitus (GDM), a common medical complication of pregnancy, is defined as "diabetes diagnosed in the second or third trimester of pregnancy that was not clearly overt diabetes prior to gestation[1]". The estimated global prevalence of GDM in 2021 was between 4.3% and 38.1%[2]. The prevalence in North America and the Caribbean is reported to be 20.7%[3]. In Mexico, the prevalence of GDM is estimated to be between 10 and 12%, though insufficient data are available to determine the national prevalence[4,5]. Pregnant women with GDM have

[1]University of Guadalajara, Health Sciences University Center, 44340 Guadalajara, Jalisco, Mexico. [2]Carlos Slim Foundation, Lago Zurich 245, Presa Falcon Building (Floor 20), Col. Ampliacion Granada, 11529 Mexico City, Miguel Hidalgo, Mexico. [3]National Institute of Public Health, Universidad 655, Santa María Ahuacatitlan, 62100 Cuernavaca, Mexico. [4]Maternal and Childhood Research Center (CIMIGEN), Tlahuac 1004, Iztapalapa, 09890 Mexico City, Mexico. [5]Ministry of Health of the State of Guanajuato, Tamazuca 4, 36000 Guanajuato, Gto, Mexico. [6]Ministry of Health of the State of Hidalgo, Fraccionamiento Puerta de Hierro, Avenida de La Mineria 103, 42080 Pachuca, Hidalgo, Mexico. [7]Department of Endocrinology, National Institute of Perinatology, Montes Urales 800, Lomas de Chapultepec, Miguel Hidalgo, 11000 Mexico City, Mexico. [8]School of Medicine, National Autonomous University of Mexico, Universidad 3004, Coyoacan, 04510 Mexico City, Mexico. ✉email: mjuarezl@fundacioncarlosslim.org

an increased risk of both short- and long-term maternal and perinatal complications[6]. Short-term complications include perinatal death, preeclampsia, and neonatal respiratory distress syndrome, among others[6,7]. Long-term complications related to GDM for both mother and infant can include impaired glucose metabolism, type 2 diabetes, cardiovascular disease/hypertension, and obesity[6,8]. Currently identified risk factors for developing GDM include maternal overweightness and obesity, later age at childbearing, previous history of GDM, family history of type 2 diabetes, ethnicity, and high fasting glucose level at the first prenatal visit (>5.6 mmol/L)[8,9]. Combinations of risk factors can also be predictive; for example, the combination of abdominal circumference >91.5 cm, fasting glucose >4.5 mmol/L, and a body mass index (BMI) >38.6 kg/m^2 greatly increases the risk of GDM[10]. Genetic risk factors for developing GDM are also being explored[11].

Screening and diagnosis of GDM occurs between weeks 24 and 28 of pregnancy; the gold standard for GDM diagnosis is an oral glucose tolerance test (OGTT)[1]. Early detection and initiation of care for GDM are key to successful interventions. Initial treatment often involves non-pharmacological interventions such as dietary modifications and increased physical activity, followed by pharmacotherapy if initial efforts do not result in normoglycemia[8]. GDM treatment reduces serious perinatal morbidity and treatment of even mild GDM can reduce the risk of fetal overgrowth, shoulder dystocia, and cesarean delivery[12,13]. Intrauterine exposure to maternal diabetes is associated with an increased risk for type 2 diabetes and obesity[14–16], although screening for GDM during pregnancy does not appear to improve childhood obesity risk[17,18]. Excessive growth of the fetal abdominal circumference, occurring between 20 and 28 weeks, precedes a GDM diagnosis, suggesting that the onset of fetal growth disorder occurs prior to the usual time of GDM screening (24 to 28 weeks)[19]. Given a high birth weight is associated with an increased risk of childhood obesity[20], it is possible that some effects of GDM begin to manifest in the fetus prior to the recommended screening and that outcomes may improve with an earlier GDM diagnosis. A study comparing outcomes in high-risk women diagnosed with GDM early in their pregnancy (prior to 24 weeks) with those diagnosed from 24 weeks found a benefit for earlier GDM diagnosis. Earlier diagnoses resulted in a significantly lower new-born composite outcome frequency (hypoglycemia, birth trauma, neonatal intensive care unit/special care nursery admission, stillbirth, neonatal death, respiratory distress, and phototherapy) compared with a later GDM diagnosis[21].

While early diagnosis of GDM is key to improving outcomes, maintaining the gold standard of universal OGTT for pregnant women is not sustainable in low-resource countries, as they often lack the infrastructure to provide universal access to facilities that can perform OGTT[22]. Additionally, there are barriers to performing the OGTT, including the inconvenience of a multi-hour appointment with multiple blood draws, difficulty in tolerating the test protocol, and the lack of understanding of the disease by healthcare workers and pregnant women[23]. GDM diagnostic capacity was reduced during the COVID-19 pandemic as OGTT testing has been difficult; several countries modified their diagnostic approach to GDM to limit long appointments[24]. Modified diagnostic approaches, such as the utilization of faster, less accurate testing, may result in underdiagnosis of GDM[25]. This lost diagnostic capacity during the COVID-19 pandemic was not limited to GDM. A study conducted in Spain found an average decline in new diagnoses of 31.1% in 2020 compared with 2019[26].

Artificial intelligence (AI) is being applied in healthcare to facilitate solutions to complex real-world problems in this current era of digital health. As such, it is important to evaluate potential new ways to utilize AI-based technologies that have the potential to change the delivery of primary healthcare and preventative medicine. Early risk stratification by prediction modelling based on AI is expected to facilitate early detection of GDM in at-risk populations that would benefit from early intervention. The ability to predict the risk of developing GDM during pregnancy, which increases the risk of complications during pregnancy, represents a useful companion to therapeutic decision-making and patient education. Models currently developed to predict GDM use clinical parameters and serum biomarkers that are not always accessible (e.g., OGTT in pregnant women is not readily available in Mexico)[27–30]. A tool for predicting the risk of developing GDM that requires less medical equipment and fewer laboratory tests is needed to facilitate early interventions for women with a high risk of developing GDM. This would be particularly useful in low- and middle-income countries, where a lack of resources and trained personnel may result in placing a lower priority on the diagnosis and care of women with GDM[31]. The remarkable performance of AI in predicting the risk for other diseases, such as coronary heart disease among patients with type 2 diabetes mellitus[32], cardiovascular disease[33], and diabetic retinopathy[34], suggests it may hold promise in applications to predict GDM risk. We aimed to develop an AI-based prediction model for risk of developing GDM among pregnant women in Mexico.

Results

Screening population. The *Cuido Mi Embarazo* (CME) study collected data for 1709 pregnant women in Mexico between April 2019 and May 2021. For this work we used information from 860 women (430 with GDM, and 430 without GDM). The sociodemographic and clinical characteristics according to GDM diagnosis are shown in Table 1. There were significant differences in the non-GDM and GDM groups for age, pregestational BMI, parity, family history of diabetes, BMI at study enrollment, random capillary glucose at study enrollment, and fasting plasma glucose measured between the 24th and 28th week of pregnancy (first OGTT measurement).

AI predictive model. Our model was named the Medición Integrada para la Detección Oportuna (MIDO) AI model for predicting gestational diabetes (MIDO GDM). The training data subset included 860 pregnant women, the validation data subset included 86, and the test data subset included 86 (corresponding to 10% of the data subset, respectively). The training of this model was used as a pilot for the MIDO GDM, which is expected to continuously improve as additional information becomes available. Table 2 shows the sensitivity and specificity of the MIDO GDM model at different cut-off points. Analysis of the risk distribution showed that the optimal cut-off point to provide an ideal balance for sensitivity and specificity was ≥ 0.40, which provided a sensitivity of

Characteristics	Non-GDM group (n = 430)	GDM group (n = 430)	P value
Age categories (n %)			
min – 20 years	59 (13.7)	31 (7.2)	
20.1 – 25 years	122 (28.4)	87 (20.2)	
25.1 – 34.9 years	203 (47.2)	223 (51.9)	0.0001
> 35 years	46 (10.7)	89 (20.7)	
Pregestational BMI (n %)			
min – 25	237 (55.1)	128 (30.0)	
25.1 – 30	130 (30.2)	184 (42.6)	
30.1 – 35	41 (9.5)	82 (19.0)	0.0001
> 35	13 (3.0)	36 (8.4)	
Parity (n %)			
Nulliparous	189 (44.4)	167 (38.6)	
Multiparous	237 (55.6)	263 (61.4)	0.101
Family history of diabetes mellitus (n %)	121 (28.1)	254 (58.8)	0.0001
Family history of hypertension	104 (24.2)	57 (13.3)	0.806
History of hypertension (n %)	1 (0.2)	1 (0.2)	0.637
Gestational week	19.5 ± 6.6	20.2 ± 6.8	0.211
Enrollment BMI (n %)			
min–25	160 (37.2)	105 (24.4)	
25.1–30	125 (29.0)	159 (36.7)	
30.1–35	47 (11.0)	79 (18.4)	0.0001
> 35	10 (2.3)	37 (8.6)	
Data not available	88 (20.5)	51 (11.9)	
Capillary casual glucose at enrollment (mg/dL)	95.4 ± 19.5	99.3 ± 17.0	0.078
OGTT fasting serum glucose (mg/dL)	77.7 ± 7.8	93.9 ± 10.9	0.0001
OGTT 1-h serum glucose (mg/dL)	116.2 ± 29.7	165.2 ± 38.6	0.0001
OGTT 2-h serum glucose (mg/dL)	99.8 ± 21.0	137.7 ± 33.8	0.0001

Table 1. Sociodemographic and clinical characteristics of the included pregnant women according to GDM status. *BMI* body mass index, *GDM* gestational diabetes mellitus, *GW* gestational week, *OGTT* oral glucose tolerance test, Continuous variables are expressed as mean ± standard deviation, Chi-square test was used to evaluate qualitative variables, student's *t* test was used to evaluate continuous variables.

Cut-off point	Sensitivity (%)	Specificity (%)	Correct classification (%)
≥ 0.06	99.77	5.08	28.91
≥ 0.23	94.88	36.28	51.02
≥ 0.40	85.12	67.40	71.85
≥ 0.53	80.70	76.47	77.53
≥ 0.81	46.05	93.51	81.57
≥ 0.90	38.14	98.44	83.27

Table 2. Sensitivity and specificity of the MIDO GDM model at different cut-off points. *MIDO GDM* Medición Integrada para la Detección Oportuna (MIDO) artificial intelligence model for predicting gestational diabetes mellitus.

85.12% and a specificity of 67.40% with a rate of correct classification of 71.85%. The artificial neuronal network (ANN) reached a precision of 84.4%, 77.9%, and 75.7% in the training, validation, and test data subsets, respectively. The model was then used to estimate the risk for all pregnant women included in the cohort (n = 1709), achieving a precision of 70.28% and a sensitivity of 83.26%.

A kernel density plot was generated using a bandwidth of 0.0579 to visualize the distribution of predicted values from the ANN model. The plot demonstrates a bimodal distribution, showing pregnant women with a lower risk of developing GDM concentrating on the left side of the plot and those with a higher risk concentrating on the right side (Fig. 1). The area under the receiver operating characteristic (ROC) curve on the test set was 0.8471 (95% confidence interval 0.8253–0.8689), indicating good accuracy of the model (Fig. 2).

External validation of the model showed an area under the ROC curve of 0.9308 with a precision of 86.00%, a specificity of 86.96%, and a sensitivity of 85.19%.
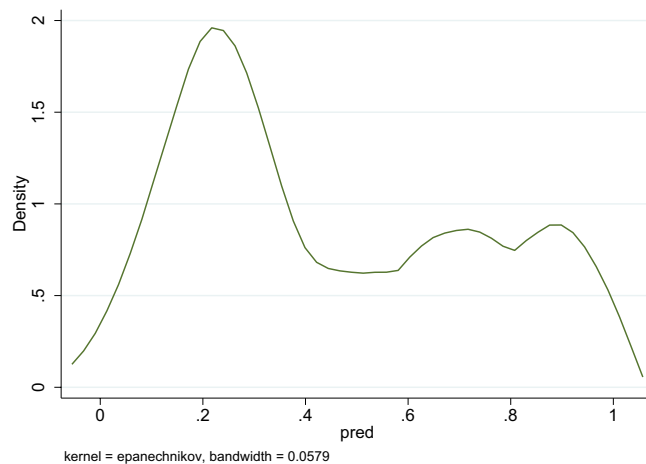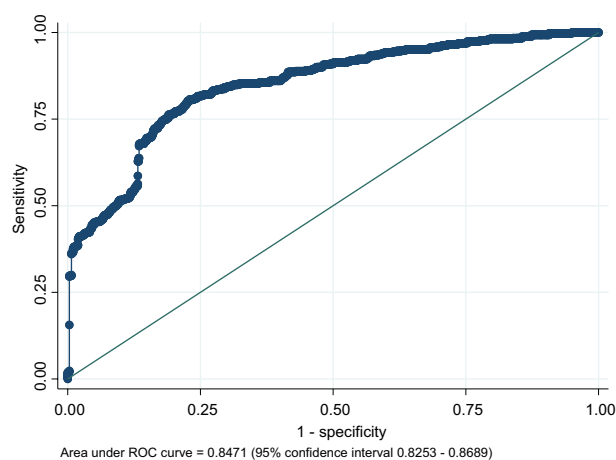
Figure 1. Kernel density plot.

Figure 2. Area under the ROC curve for the MIDO GDM artificial neural network. *MIDO GDM* Medición Integrada para la Detección Oportuna (MIDO) artificial intelligence model for predicting gestational diabetes mellitus, *ROC* receiver operating characteristic.

Discussion

The current era of digital health provides an opportunity to change traditional models of primary healthcare. We have developed a novel digital health solution, MIDO GDM, which represents the first predictive model of GDM for pregnant Mexican women that is accurate in the first trimester of pregnancy, providing an opportunity for early detection that can be utilized prior to the use of the OGTT at 24 weeks. This approach provides an opportunity to offer personalized strategies to prevent the development of diabetes in pregnancy. Using MIDO GDM, pregnant women at risk of developing GDM can be identified in the first trimester of pregnancy with a group of easy-to-obtain variables. Thus, it would be possible to establish the risk of developing GDM in the early stages of pregnancy and to promote personalized preventive actions and appropriate treatment in a timely manner.

In our study, we identified nine variables that enabled the model to determine risk for developing GDM with a high level of precision, thus allowing us to propose an ANN that is simple to use and grants healthcare providers the opportunity to prioritize which patients receive preventative advice and treatment over time. Studies of early interventions (prior to pregnancy week 24) in women who are not yet diagnosed with GDM have reported promising results[35]. Among Finnish women who were < 20 weeks pregnant and at a high risk of developing GDM, those who received counselling on diet, physical activity, and weight control, and those who met with a dietician, had a significantly lower incidence of GDM compared with women who did not receive these interventions ($p = 0.044$)[36]. Another study that enrolled women with a BMI > 25 kg/m^2 at their first trimester found that those enrolled in a Therapeutic Lifestyle Changes program, which included diet and mild physical activity, had a significantly lower incidence of GDM compared with those who received no intervention ($p = 0.009$)[37]. These studies support a potential benefit to the early identification of pregnant women at high risk for developing GDM. We propose that the MIDO GDM model be used ubiquitously throughout Mexico to improve the timing and quality of interventions for the prevention of GDM, a disease that is poorly diagnosed in the population.

We expect that MIDO GDM will play an instrumental role in changing the way primary care and preventative medicine are delivered. As an ANN, MIDO GDM will continue to learn as it is utilized as a part of the digital health platforms currently in use in Mexico. The MIDO GDM is expected to improve healthcare guidance by physicians who use MIDO Integral and to provide risk assessments and personalized health and lifestyle recommendations for patients who access the Mi Salud Integral app.

The success of a predictive model lies in its ability to predict, learn, and generate improvements that provide real-world value. As such, MIDO GDM will be incorporated within the MIDO healthcare strategy[38,39], which is a series of interconnected digital platforms that enable healthcare professionals to proactively prevent obesity, type 2 diabetes mellitus, hypertension, and dyslipidemia at the community level (where laboratory tests are available)[39]. This ecosystem is known as MIDO Integral, and aims to provide timely and quality healthcare that is personalized to each person's life course. For pregnant women, the MIDO Pregnancy system focuses on providing screening for diabetes, hypertension, weight gain during pregnancy, and urinary tract infections.

A diagram of the incorporation of the MIDO GDM model into the Mexican healthcare system is shown in Fig. 3. It is important that AI-based models do not merely remain conceptual but are integrated into healthcare systems. Because MIDO GDM is an ANN, the model will continuously learn as it is being used among the Mexican population within a functioning digital health ecosystem (MIDO Integral Ecosystem)[40,41]. Both healthcare professionals (via MIDO Integral) and patients (from home via the Mi Salud Integral app[42,43]) will have access to risk assessments and personalized health and lifestyle recommendations. This information will inform healthcare guidance and patient self-care.

MIDO Pregnancy is currently utilized across 388 primary healthcare centers in 32 states in Mexico, offering personalized recommendations to both healthcare professionals and patients based on their screening test results. Recommendations are made on the basis of decision trees and are given at the end of the assessment based on the results of the screening. Screening is performed at every visit to the health center and therefore our predictive model can be used to detect the risk of GDM from the very first prenatal visit in order to provide personalized GDM-prevention strategies as early as possible. MIDO GDM will also be able to offer and receive information from the mobile application "Mi Salud Integral". We expect that the application of this strategy will benefit primary healthcare and, in a larger scope, change public policy and shift the traditional primary healthcare models in Mexico and other countries.

The MIDO GDM model is simple and easy to implement even in low-resource settings, thus enhancing the practice of precision medicine. There is increasing evidence that digital health tools have the potential to improve patient health in a variety of clinical areas for those living in low- and middle-income countries, though it is widely recognized that the opportunities to use digital health tools are underutilized[40,44–46]. Under these settings, there is limited per capita spending on healthcare services, leaving many people without access to healthcare[40].

This study had several strengths. First, we used variables with commonly and easily collected data, which made it feasible to collect sufficient data to develop this predictive model. The utilization of easily accessible data means that our model can be applied in low-resource settings, where frequent laboratory testing is not always feasible due to resource limitations. Second, we developed this model using data collected as part of a
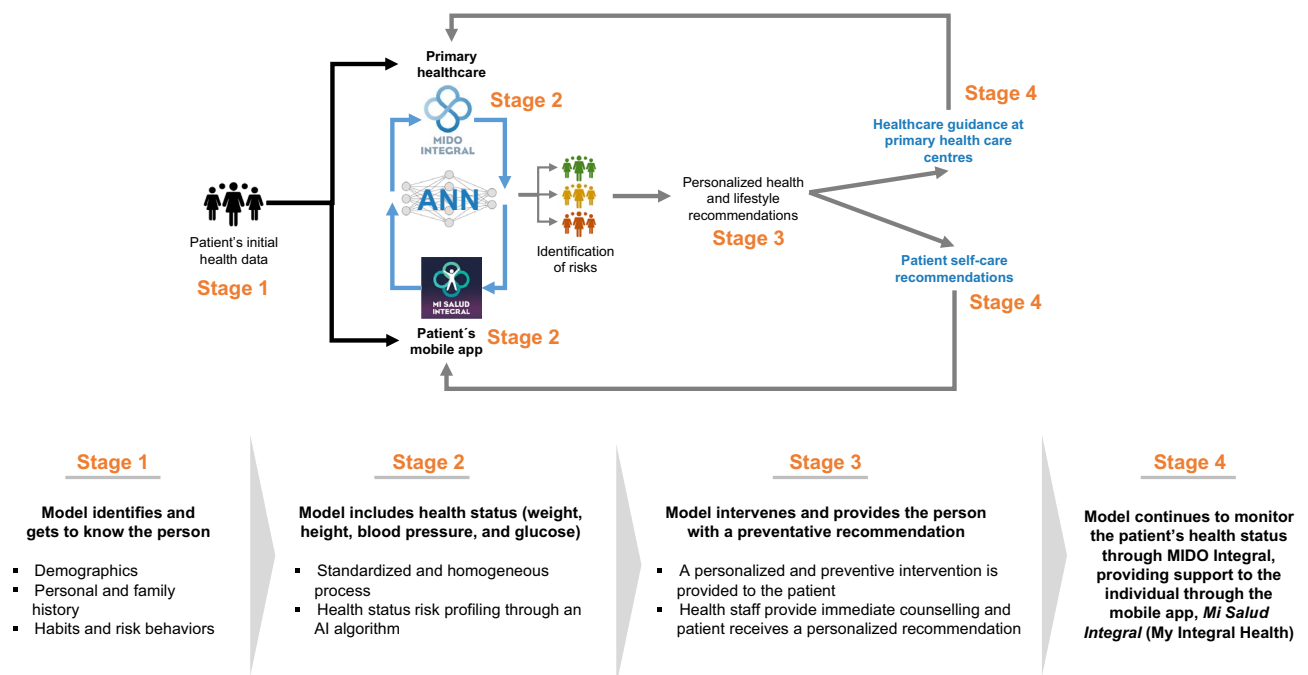

Figure 3. MIDO GDM as part of a proactive digital strategy for the detection of disease risks and disease. *AI—artificial intelligence, ANN—artificial neuronal network, MIDO GDM—Medición Integrada para la Detección Oportuna (MIDO) artificial intelligence model for predicting gestational diabetes mellitus.*

multicenter cohort study, in which all pregnant women received OGTT according to both national and international standards. Third, we used ANNs to develop our model, which are not only able to discover patterns and relationships in large amounts of data, they are also able to be trained to produce the desired input–output relationship by running the parameters to adjust the network behavior[47]. This gives them an advantage over conventional computing methods. Finally, our ANN model performs similarly or better than several models recently developed to provide early prediction of GDM in pregnant women. Our model had an AUC, sensitivity, and specificity of 0.8471, 83.26%, and 70.28%, respectively. A predictive model using a deep neuronal network reported respective values of 0.80, 63%, and 82%[46]. Similar values were also observed when applying a logistic regression model (AUC, 0.64–0.77; sensitivity, 58–69%; specificity, 53–86%)[48–50] or a random forest model (AUC, 0.777; sensitivity, 65.1%; specificity, 81.3%)[49].

This study also had several limitations. The model was based on a small set of data from Mexican women; therefore, the generalizability of the model to the entire population of Mexican women and other ethnic populations is unknown. Given the promising precision and sensitivity of the model, studies further validating this model using larger datasets are warranted. Aside from the limitations of the study, the model itself also has several limitations. First, some variables may be unavailable for some women; however, the effects of this may be minimized by imputation of the missing information. For example, although over 50% of women had missing values for glucose level at the time of their first pregnancy consultation, this variable is of great relevance for prediction. Unfortunately, in Mexico there are several barriers to routinely providing glucose testing at the first pregnancy consultation; thus, we were not surprised to see that data for this parameter were missing for many pregnant women. Considering these factors, we considered it appropriate to retain this variable by imputation of the missing information. Second, although false positives are likely, a positive readout will trigger advice on non-pharmaceutical preventative strategies that benefit all pregnant women rather than a pharmaceutical intervention. Third, tools such as MIDO GDM may not be easy for healthcare professionals to adopt. Finally, the usefulness of this tool will be dependent upon whether pregnant women successfully adopt preventative measures when advised.

In conclusion, we utilized ANNs to build a model that could predict the risk of developing GDM using variables for which data is commonly and easily collected. In building this model, we aimed to alleviate some of the uncertainty in the therapeutic decision-making process for pregnant women in Mexico and to assist physicians in making decisions related to clinical care. Multiple ANN algorithms were used to develop this model, which we expect can be used to identify at-risk pregnant women during the early stages of pregnancy, despite the limited availability of OGTT in low-resource settings. Early detection of GDM is critical for initiating timely preventative measures and successful interventions. Furthermore, it is hoped that this AI-based neural network will potentially play an instrumental role in changing the way primary care and preventative medicine is delivered.

Methods

This study was conducted to build and train a neural network AI algorithm to predict the risk of developing GDM in Mexican women using socio-demographic and clinical variables, which were considered relevant for the detection of GDM risk. All women provided written informed consent prior to participation in the CME study. The study protocol was approved by the Research and Ethics Committees of the Secretary of Health of Hidalgo State (FSSA2018076) and the Secretary of Health of Guanajuato State (CONBIOETICA-11-CEI-003–20,190,704). The study was conducted in accordance with the Declaration of Helsinki, the International Conference on Harmonization guidelines for Good Clinical Practice, and applicable local laws and regulations. The study was registered at researchregistry.com (researchregistry7405).

Data sources. The study population was derived from a prospective multicenter cohort study called CME [researchregistry7405]; this included data for 1709 pregnant women in Mexico, collected between April 2019 and May 2021. The CME study aimed to achieve detection of GDM in six primary healthcare centers in Mexico and to provide appropriate and timely care for GDM. Data were collected using standardized questionnaires for socio-demographic characteristics (address, type of medical affiliation, marital status), medical history (family and personal history of pathologies, blood type, gynecological history, medications used), regular medical visits (maternal weight, blood pressure, urine test strip [glucose/protein/pH/blood/ketones/bilirubin/nitrite/urobilinogen/specific gravity/leukocytes], alarm signs), laboratory data (hemoglobin, hematocrit, serum platelet level, serum glucose level, lipid profile, urine pregnancy test at each trimester), pregnancy resolution (date pregnancy ended, delivery method, number of newborns, complications during delivery), and management of self-monitoring for GDM (data related to at-home glucose monitoring using a capillary glucometer provided to participants of the cohort study). All participants in this cohort were referred for GDM testing. Clinical evaluation, including a 2-h 75-g OGTT that was performed between weeks 24 and 28 of their pregnancy in accordance with International Association of Diabetes and Pregnancy Study Group criteria[51], was used to establish a GDM diagnosis, which was given if any single value met or exceeded normal glucose threshold values specified as normal for pregnant Mexican women (fasting value of 92 mg/dL [5.11 mmol/L]; a 1-h value of 180 mg/dL [10.0 mmol/L]; or a 2-h value of 153 mg/dL [8.49 mmol/L])[1,51]. In this analysis, we included pregnant women without type 2 diabetes and pregnant women who were at < 36 weeks gestation from the CME cohort. Furthermore, those diagnosed with pregestational diabetes, had multiple pregnancies, or had a previous chronic disease that required their pregnancy to be monitored by secondary care were excluded.

Predictive model development. The use of AI, and ANNs specifically, has been investigated in a variety of healthcare applications, including clinical diagnosis, disease prediction and detection, risk stratification, and image analysis and interpretation[52–55]. More recently, ANNs are being used to inform healthcare management

decisions with a goal of moving towards value-based care[55]. We used an ANN approach to develop our predictive model, which has advantages over traditional statistical methods, including the ability to work with non-structured and missing input data, the ability to identify complex non-linear relationships between predictive variables, parallel data processing, and the ability to learn[47,56]. A "weight" analysis can be used as a sensitivity analysis to explain relationships between input and output variables[57]. This type of analysis computes, quantitatively, the strength of connections between input and output factors. It is expected that this characteristic of the ANN approach will result in a high degree of predictive accuracy. Thus, ANN is ideal for developing a model to predict the risk of developing GDM.

We took the following steps to develop our ANN model: (i) data pre-processing, (ii) building and training the model, (iii) simplifying the model, adding dropout and weight regularization to address overfitting, and (iv) hyperparameter tuning. First, all candidate predictive variables that were selected after the systematic review were included to build a base model. Then, we conducted a machine-learning variable selection method to select variables that ensured better model discrimination to create an efficient approach for clinical practice with fewer redundant variables. We evaluated variables with predictive power that could be easily collected by questionnaire and did not require laboratory measurements. This strategy was employed to ensure the model could be used in situations where laboratory measurements were not readily available. We selected nine variables for GDM prediction after reviewing 28 published studies that reported findings from cohort and case–control studies, diagnosis or assessments of GDM, or development of predictive or risk models[27–30,58–81] (listed and summarized in Table S1 and Figure S1) and in accordance with the information available for the CME study. The selected variables included maternal age (years); family history of type 2 diabetes (yes/no); previous diagnosis of hypertension (yes/no); BMI (kg/m^2); gestational week at first prenatal visit; parity (number); birth weight of last child (grams); type of capillary glucose measurement at first prenatal visit (random/fasting); and capillary blood glucose level at first prenatal visit (mg/dL).

Missing data for each variable were imputed with the average, bordering, or null value observed in a typical pregnancy for a Mexican woman based on information provided by the Mexico Births Information Subsystem (SINAC) for the year 2020 (Table 3)[40]. All variables were standardized using the mean and standard deviation to be included into the model. The dataset for model building was first balanced to equalize the number of women with and without GDM, and then it was divided into subsets for training (80%), validation (10%), and testing (10%).

Figure 4 shows the ANN architecture of the model. The model had a single output node, defined as the GDM diagnosis (yes/no), which was modelled using an ANN approach with the following hyper-parametrization: three hidden layers of 10, 8, and 6 nodes; Rectified Linear Activation as the hidden layer activation function; sigmoid or logistics as the output activation function; and a binary cross-entropy as loss of function to measure how far away from the true value the prediction was to allow for tuning of the parameters based on this value. Next, we used Adam, an adaptive learning rate optimization algorithm and set the number of epochs (i.e., the number of times the entire training dataset is shown to the network during training) as 100 and the batch size (i.e., the number of sub-samples given to the network after the parameter update occurs) as 50 using the Early Stopping technique. Finally, we assessed the accuracy of the model to predict GDM.

After obtaining the risk predictions, we performed an analysis of the risk distribution to define the optimal cut points that would provide an ideal balance for sensitivity and specificity. Analyses were performed using the Pandas, Numpy, and Keras (backend on tensorflow) modules in Python 3.6 (Python Software Foundation, Fredericksburg, VA, USA).

External validation of the model. The model was externally validated using data collected during routine health care provided to pregnant women in Mexico City who visited the Health Units of the Maternal and Infant Research Center of the Birth Study Group (CIMIGEN [acronym in Spanish])[82]. Data were collected from the randomly selected records of 27 pregnant women with GDM and 23 pregnant controls without GDM who had their first medical appointment at ≤ 32 weeks of pregnancy. Data related to the variables used in the model were retrieved, including maternal age, family history of type 2 diabetes, previous diagnosis of hypertension, pregestational BMI, week of pregnancy at prenatal visit, parity, birth weight of last child, type of capillary glucose measurement at first prenatal visit (random/fasting), and capillary blood glucose level at prenatal visit. The

Input variable	Value†	Missing values (% of total)
Age, years	26	11.8
Family history of diabetes‡	0	0.2
Family history of hypertension‡	0.5	12.3
BMI before pregnancy, kg/m^2	25	13.9
Gestational week at time of first pregnancy consultation	18	11.8
Number of gestations	1	0.5
Birth weight of the last child, g	3100	43.3
Type of glucose test§	0.5	64.1
Glucose level at time of first pregnancy consultation, mg/dL	92	61.5

Table 3. Values used to impute missing data. *BMI* body mass index. †Average value reported for a typical pregnancy in Mexico. ‡No = 0, yes = 1. §Fasting = 1 or random = 0.
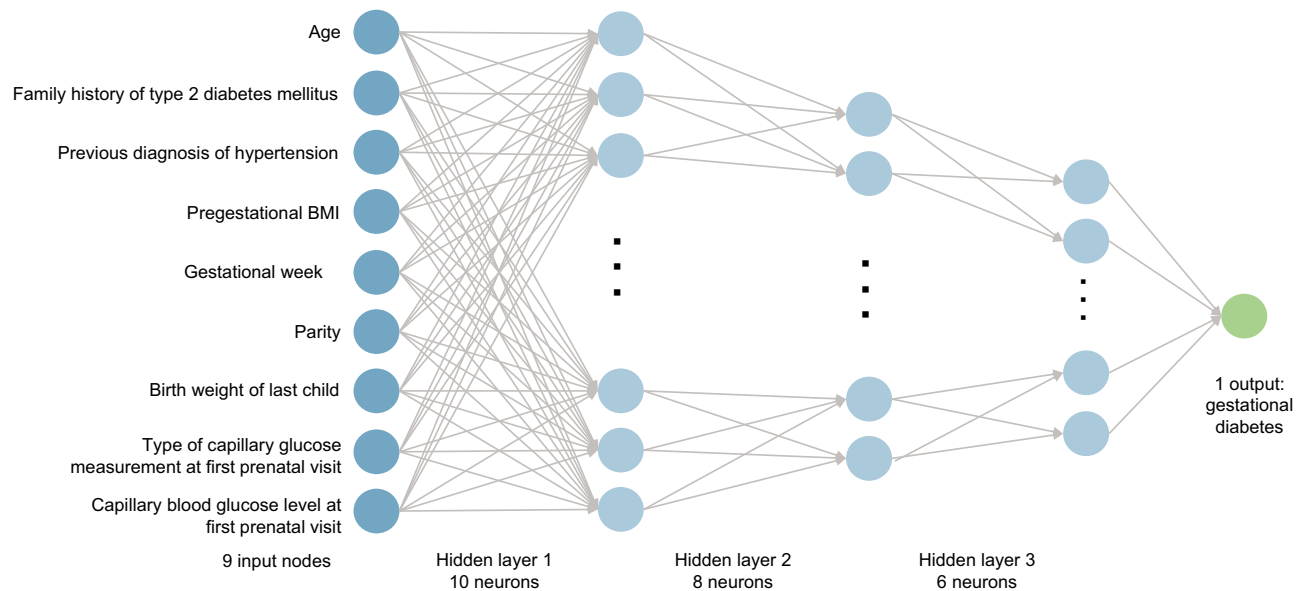
Figure 4. Artificial neural network architecture. *BMI*—body mass index, *GDM*—gestational diabetes mellitus.

selected pregnant women were confirmed to not be part of the cohort used to develop the predictive model. The resulting database (Supplementary Data S1) was carefully reviewed, and the completeness and plausibility of the dataset were confirmed prior to running the model.

Data availability

All data collected for this study will be made available upon reasonable request to the corresponding author.

References

1. Classification and diagnosis of diabetes. Standards of medical care in diabetes-2021. *Diabetes Care* **44**(Suppl 1), S15–S33. https://doi.org/10.2337/dc21-S002 (2021).
2. International diabetes federation. https://diabetesatlas.org/atlas/tenth-edition/.
3. International diabetes federation. https://diabetesatlas.org/data/en/indicators/14/.
4. Reyes-Muñoz, E., Parra, A., Castillo-Mora, A. & Ortega-González, C. Effect of the diagnostic criteria of the international association of diabetes and pregnancy study groups on the prevalence of gestational diabetes mellitus in urban Mexican women: A cross-sectional study. *Endocr. Pract.* **18**(2), 146–151. https://doi.org/10.4158/EP11167.OR (2012).
5. Centro nacional de equidad de género y salud reproductiva. http://cnegsr.salud.gob.mx/contenidos/descargas/SMP/LineamientoDiabetesyEmbarazo.pdf.
6. Management of diabetes in pregnancy. Standards of medical care in diabetes-2021. *Diabetes Care* **44**(Suppl 1), S200–S210. https://doi.org/10.2337/dc21-S014 (2021).
7. Schmidt, M. I. *et al.* Gestational diabetes mellitus diagnosed with a 2-h 75-g oral glucose tolerance test and adverse pregnancy outcomes. *Diabetes Care* **24**(7), 1151–1155. https://doi.org/10.2337/diacare.24.7.1151 (2001).
8. McIntyre, H. D. *et al.* Gestational diabetes mellitus. *Nat. Rev. Dis. Primers* **5**(1), 47. https://doi.org/10.1038/s41572-019-0098-8 (2019).
9. Popova, P. *et al.* Fasting glycemia at the first prenatal visit and pregnancy outcomes in Russian women. *Minerva Endocrinol.* **41**(4), 477–485 (2016).
10. Popova, P. V. *et al.* The new combination of risk factors determining a high risk of gestational diabetes mellitus. *Minerva Endocrinol.* **40**(4), 239–247 (2015).
11. Popova, P. V. *et al.* Association of common genetic risk variants with gestational diabetes mellitus and their role in GDM prediction. *Front. Endocrinol. Lausanne* **12**, 628582 (2021).
12. Crowther, C. A. *et al.* Effect of treatment of gestational diabetes mellitus on pregnancy outcomes. *N. Engl. J. Med.* **352**(24), 2477–2486. https://doi.org/10.1056/NEJMoa042973 (2005).
13. Landon, M. B. *et al.* A multicenter, randomized trial of treatment for mild gestational diabetes. *N. Engl. J. Med.* **361**(14), 1339–1348. https://doi.org/10.1056/NEJMoa0902430 (2009).
14. Dabelea, D. *et al.* Intrauterine exposure to diabetes conveys risks for type 2 diabetes and obesity: A study of discordant sibships. *Diabetes* **49**(12), 2208–2211. https://doi.org/10.2337/diabetes.49.12.2208 (2000).
15. Dabelea, D. *et al.* Association of intrauterine exposure to maternal diabetes and obesity with type 2 diabetes in youth: The SEARCH case-control study. *Diabetes Care* **31**(7), 1422–1426. https://doi.org/10.2337/dc07-2417 (2008).
16. Hartling, L. *et al.* Screening and diagnosing gestational diabetes mellitus. *Evid. Rep. Technol. Assess Full Rep.* **210**, 1–327 (2012).
17. Gillman, M. W. *et al.* Effect of treatment of gestational diabetes mellitus on obesity in the next generation. *Diabetes Care* **33**(5), 964–968. https://doi.org/10.2337/dc09-1810 (2010).
18. Landon, M. B. *et al.* Mild gestational diabetes mellitus and long-term child health. *Diabetes Care* **38**(3), 445–452. https://doi.org/10.2337/dc14-2159 (2015).
19. Sovio, U., Murphy, H. R. & Smith, G. C. Accelerated fetal growth prior to diagnosis of gestational diabetes mellitus: A prospective cohort study of nulliparous women. *Diabetes Care* **39**(6), 982–987. https://doi.org/10.2337/dc16-0160 (2016).

20. Qiao, Y. *et al.* Birth weight and childhood obesity: A 12-country study. *Int. J. Obes. Suppl.* **5**(Suppl 2), S74-79. https://doi.org/10.1038/ijosup.2015.23 (2015).

21. Clarke, E., Cade, T. J. & Brennecke, S. Early pregnancy screening for women at high-risk of GDM results in reduced neonatal morbidity and similar maternal outcomes to routine screening. *J. Pregnancy* **2020**, 9083264. https://doi.org/10.1155/2020/9083264 (2020).

22. Putoto, G. *et al.* A simplified diagnostic work-up for the detection of gestational diabetes mellitus in low resources settings: Achievements and challenges. *Arch. Gynecol. Obstet.* **302**(5), 1127–1134. https://doi.org/10.1007/s00404-020-05708-x (2020).

23. Lachmann, E. H. *et al.* Barriers to completing oral glucose tolerance testing in women at risk of gestational diabetes. *Diabet. Med.* **37**(9), 1482–1489. https://doi.org/10.1111/dme.14292 (2020).

24. McIntyre, H. D. *et al.* Testing for gestational diabetes during the COVID-19 pandemic. An evaluation of proposed protocols for the United Kingdom, Canada and Australia. *Diabetes Res. Clin. Pract.* **167**, 108353. https://doi.org/10.1016/j.diabres.2020.108353 (2020).

25. van Gemert, T. E., Moses, R. G., Pape, A. V. & Morris, G. J. Gestational diabetes mellitus testing in the COVID-19 pandemic: The problems with simplifying the diagnostic process. *Aust. N. Z. J. Obstet. Gynaecol.* **60**(5), 671–674. https://doi.org/10.1111/ajo.13203 (2020).

26. Pifarre, I. A. H. *et al.* Missing diagnoses during the COVID-19 Pandemic: A year in review. *Int. J. Environ. Res. Public Health* **18**(10), 5335. https://doi.org/10.3390/ijerph18105335 (2021).

27. Teede, H. J., Harrison, C. L., Teh, W. T., Paul, E. & Allan, C. A. Gestational diabetes: Development of an early risk prediction tool to facilitate opportunities for prevention. *Aust. N. Z. J. Obstet. Gynaecol.* **51**(6), 499–504. https://doi.org/10.1111/j.1479-828X.2011.01356.x (2011).

28. Nanda, S., Savvidou, M., Syngelaki, A., Akolekar, R. & Nicolaides, K. H. Prediction of gestational diabetes mellitus by maternal factors and biomarkers at 11 to 13 weeks. *Prenat. Diagn.* **31**(2), 135–141. https://doi.org/10.1002/pd.2636 (2011).

29. van Leeuwen, M. *et al.* Estimating the risk of gestational diabetes mellitus: A clinical prediction model based on patient characteristics and medical history. *BJOG* **117**(1), 69–75. https://doi.org/10.1111/j.1471-0528.2009.02425.x (2010).

30. Syngelaki, A. *et al.* First-trimester screening for gestational diabetes mellitus based on maternal characteristics and history. *Fetal Diagn. Ther.* **38**(1), 14–21. https://doi.org/10.1159/000369970 (2015).

31. Goldenberg, R. L., McClure, E. M., Harrison, M. S. & Miodovnik, M. Diabetes during pregnancy in low- and middle-income countries. *Am. J. Perinatol.* **33**(13), 1227–1235. https://doi.org/10.1055/s-0036-1584152 (2016).

32. Fan, R. *et al.* AI-based prediction for the risk of coronary heart disease among patients with type 2 diabetes mellitus. *Sci. Rep.* **10**(1), 14457. https://doi.org/10.1038/s41598-020-71321-2 (2020).

33. Alaa, A. M., Bolton, T., Di Angelantonio, E., Rudd, J. H. F. & van der Schaar, M. Cardiovascular disease risk prediction using automated machine learning: A prospective study of 423,604 UK Biobank participants. *PLoS One* **14**(5), e0213653. https://doi.org/10.1371/journal.pone.0213653 (2019).

34. Gunasekeran, D. V., Ting, D. S. W., Tan, G. S. W. & Wong, T. Y. Artificial intelligence for diabetic retinopathy screening, prediction and management. *Curr. Opin. Ophthalmol.* **31**(5), 357–365. https://doi.org/10.1097/ICU.0000000000000693 (2020).

35. Zhang, C., Rawal, S. & Chong, Y. S. Risk factors for gestational diabetes: Is prevention possible?. *Diabetologia* **59**(7), 1385–1390. https://doi.org/10.1007/s00125-016-3979-3 (2016).

36. Koivusalo, S. B. *et al.* Gestational diabetes mellitus can be prevented by lifestyle intervention: The finnish gestational diabetes prevention study (RADIEL): A randomized controlled trial. *Diabetes Care* **39**(1), 24–30. https://doi.org/10.2337/dc15-0511 (2016).

37. Petrella, E. *et al.* Gestational weight gain in overweight and obese women enrolled in a healthy lifestyle and eating habits program. *J. Matern. Fetal Neonatal Med.* **27**(13), 1348–1352. https://doi.org/10.3109/14767058.2013.858318 (2014).

38. Tapia-Conyer, R., Gallardo-Rincon, H. & Saucedo-Martinez, R. Casalud: Innovations for the transformation of health systems. *Gac. Med. Mex.* **148**(6), 567–571 (2012).

39. GallardoRincon, H. *et al.* Integrated Measurement for Early Detection (MIDO) as a digital strategy for timely assessment of non-communicable disease profiles and factors associated with unawareness and control: A retrospective observational study in primary healthcare facilities in Mexico. *BMJ Open* **11**(9), e049836. https://doi.org/10.1136/bmjopen-2021-049836 (2021).

40. Huckvale, K., Wang, C. J., Majeed, A. & Car, J. Digital health at fifteen: More human (more needed). *BMC Med.* **17**(1), 62. https://doi.org/10.1186/s12916-019-1302-0 (2019).

41. Mido integral. https://midointegral.com/ (2022).

42. MI SALUD INTEGRAL. https://apps.apple.com/co/app/mi-salud-integral/id1524705484 (2021).

43. MI SALUD INTEGRAL. https://play.google.com/store/apps/details?id=com.fundacionslim.micoachsalud&hl=es_MX&gl=US (2021).

44. Gibson, D. G., Tamrat, T. & Mehl, G. The state of digital interventions for demand generation in low- and middle-income countries: Considerations, emerging approaches, and research gaps. *Glob Health Sci. Pract.* **6**(Suppl 1), S49–S60. https://doi.org/10.9745/GHSP-D-18-00165 (2018).

45. Kallander, K. *et al.* Mobile health (mHealth) approaches and lessons for increased performance and retention of community health workers in low- and middle-income countries: A review. *J. Med. Internet Res.* **15**(1), e17. https://doi.org/10.2196/jmir.2130 (2013).

46. Naslund, J. A. *et al.* Digital technology for treating and preventing mental disorders in low-income and middle-income countries: A narrative review of the literature. *Lancet Psychiat.* **4**(6), 486–500. https://doi.org/10.1016/S2215-0366(17)30096-2 (2017).

47. Liao, S.-H. & Wen, C.-H. Artificial neural networks classification and clustering of methodologies and applications—Literature analysis from 1995 to 2005. *Expert Syst. Appl.* **32**, 1–11 (2007).

48. Wu, Y. T. *et al.* Early prediction of gestational diabetes mellitus in the Chinese population via advanced machine learning. *J. Clin. Endocrinol. Metab.* **106**(3), e1191–e1205. https://doi.org/10.1210/clinem/dgaa899 (2021).

49. Wang, J. *et al.* An early model to predict the risk of gestational diabetes mellitus in the absence of blood examination indexes: Application in primary health care centres. *BMC Pregnancy Childbirth* **21**(1), 814. https://doi.org/10.1186/s12884-021-04295-2 (2021).

50. Nombo, A. P., Mwanri, A. W., Brouwer-Brolsma, E. M., Ramaiya, K. L. & Feskens, E. J. M. Gestational diabetes mellitus risk score: A practical tool to predict gestational diabetes mellitus risk in Tanzania. *Diabetes Res. Clin. Pract.* **145**, 130–137. https://doi.org/10.1016/j.diabres.2018.05.001 (2018).

51. Metzger, B. E. *et al.* International association of diabetes and pregnancy study groups recommendations on the diagnosis and classification of hyperglycemia in pregnancy. *Diabetes Care* **33**(3), 676–682. https://doi.org/10.2337/dc09-1848 (2010).

52. Checcucci, E. *et al.* Artificial intelligence and neural networks in urology: current clinical applications. *Minerva Urol. Nefrol.* **72**(1), 49–57. https://doi.org/10.23736/S0393-2249.19.03613-0 (2020).

53. Chakraborty, S., Aich, S. & Kim, H. C. 3D textural, morphological and statistical analysis of voxel of interests in 3T MRI scans for the detection of Parkinson's disease using artificial neural networks. *Healthc. Basel* **8**(1), 34. https://doi.org/10.3390/healthcare8010034 (2020).

54. Isma'eel, H. A. *et al.* Artificial neural network modeling enhances risk stratification and can reduce downstream testing for patients with suspected acute coronary syndromes, negative cardiac biomarkers, and normal ECGs. *Int. J. Cardiovasc. Imaging* **32**(4), 687–696. https://doi.org/10.1007/s10554-015-0821-9 (2016).

55. Shahid, N., Rappon, T. & Berta, W. Applications of artificial neural networks in health care organizational decision-making: A scoping review. *PLoS One* **14**(2), e0212356. https://doi.org/10.1371/journal.pone.0212356 (2019).

56. Schalkoff, R. J. *Artificial Neural Networks* (McGraw-Hill, London, 1997).
57. Garson, G. D. Interpreting neural network connection weights. *Artif. Intell. Expert* **6**, 47–51 (1991).
58. Naylor, C. D., Sermer, M., Chen, E. & Farine, D. Selective screening for gestational diabetes mellitus. Toronto trihospital gestational diabetes project investigators. *N. Engl. J. Med.* **337**(22), 1591–1596. https://doi.org/10.1056/NEJM199711273372204 (1997).
59. Caliskan, E., Kayikcioglu, F., Ozturk, N., Koc, S. & Haberal, A. A population-based risk factor scoring will decrease unnecessary testing for the diagnosis of gestational diabetes mellitus. *Acta Obstet. Gynecol. Scand.* **83**(6), 524–530. https://doi.org/10.1111/j.0001-6349.2004.00389.x (2004).
60. Hedderson, M. M., Darbinian, J. A., Quesenberry, C. P. & Ferrara, A. Pregravid cardiometabolic risk profile and risk for gestational diabetes mellitus. *Am. J. Obstet. Gynecol.* **205**(1), e51-57. https://doi.org/10.1016/j.ajog.2011.03.037 (2011).
61. Gobl, C. S. *et al.* A two-step screening algorithm including fasting plasma glucose measurement and a risk estimation model is an accurate strategy for detecting gestational diabetes mellitus. *Diabetologia* **55**(12), 3173–3181. https://doi.org/10.1007/s00125-012-2726-7 (2012).
62. Trujillo, J., Vigo, A., Reichelt, A., Duncan, B. B. & Schmidt, M. I. Fasting plasma glucose to avoid a full OGTT in the diagnosis of gestational diabetes. *Diabetes Res. Clin. Pract.* **105**(3), 322–326. https://doi.org/10.1016/j.diabres.2014.06.001 (2014).
63. Fong, A., Serra, A. E., Gabby, L., Wing, D. A. & Berkowitz, K. M. Use of hemoglobin A1c as an early predictor of gestational diabetes mellitus. *Am. J. Obstet. Gynecol.* **211**(6), e641-647. https://doi.org/10.1016/j.ajog.2014.06.016 (2014).
64. Harrison, C. L., Lombard, C. B., East, C., Boyle, J. & Teede, H. J. Risk stratification in early pregnancy for women at increased risk of gestational diabetes. *Diabetes Res. Clin. Pract.* **107**(1), 61–68. https://doi.org/10.1016/j.diabres.2014.09.006 (2015).
65. Theriault, S., Forest, J. C., Masse, J. & Giguere, Y. Validation of early risk-prediction models for gestational diabetes based on clinical characteristics. *Diabetes Res. Clin. Pract.* **103**(3), 419–425. https://doi.org/10.1016/j.diabres.2013.12.009 (2014).
66. Berggren, E. K. *et al.* Variations in resting energy expenditure: Impact on gestational weight gain. *Am. J. Obstet. Gynecol.* **217**(4), 445.e1-445.e6. https://doi.org/10.1016/j.ajog.2017.05.054 (2017).
67. Capula, C. *et al.* A new predictive tool for the early risk assessment of gestational diabetes mellitus. *Prim. Care Diabetes* **10**(5), 315–323. https://doi.org/10.1016/j.pcd.2016.05.004 (2016).
68. Sweeting, A. N., Ross, G. P., Hyett, J. & Wong, J. Gestational diabetes in the first trimester: Is early testing justified?. *Lancet Diabetes Endocrinol.* **5**(8), 571–573. https://doi.org/10.1016/S2213-8587(17)30066-9 (2017).
69. Bozkurt, L. *et al.* Adiponectin and leptin at early pregnancy: Association to actual glucose disposal and risk for GDM-A prospective cohort study. *Int. J. Endocrinol.* **2018**, 5463762. https://doi.org/10.1155/2018/5463762 (2018).
70. Hinkle, S. N., Tsai, M. Y., Rawal, S., Albert, P. S. & Zhang, C. HbA1c measured in the first trimester of pregnancy and the association with gestational diabetes. *Sci. Rep.* **8**(1), 12249. https://doi.org/10.1038/s41598-018-30833-8 (2018).
71. Sweeting, A. N. *et al.* A first trimester prediction model for gestational diabetes utilizing aneuploidy and pre-eclampsia screening markers. *J. Matern Fetal Neonatal Med.* **31**(16), 2122–2130. https://doi.org/10.1080/14767058.2017.1336759 (2018).
72. Balani, J., Hyer, S. L., Shehata, H. & Mohareb, F. Visceral fat mass as a novel risk factor for predicting gestational diabetes in obese pregnant women. *Obstet. Med.* **11**(3), 121–125. https://doi.org/10.1177/1753495X17754149 (2018).
73. Pezeshki, B., Chiti, H., Arasteh, P. & Mazloomzadeh, S. Early screening of gestational diabetes mellitus using hemoglobin A1C: Revising current screening guidelines. *Casp. J. Intern. Med.* **10**(1), 16–24. https://doi.org/10.22088/cjim.10.1.16 (2019).
74. Zheng, T. *et al.* A simple model to predict risk of gestational diabetes mellitus from 8 to 20 weeks of gestation in Chinese women. *BMC Pregnancy Childbirth* **19**(1), 252. https://doi.org/10.1186/s12884-019-2374-8 (2019).
75. Zhu, Y., Hedderson, M. M., Quesenberry, C. P., Feng, J. & Ferrara, A. Central obesity increases the risk of gestational diabetes partially through increasing insulin resistance. *Obes. Silver Spring* **27**(1), 152–160. https://doi.org/10.1002/oby.22339 (2019).
76. Falcone, V. *et al.* Early assessment of the risk for gestational diabetes mellitus: Can fasting parameters of glucose metabolism contribute to risk prediction?. *Diabetes Metab. J.* **43**(6), 785–793. https://doi.org/10.4093/dmj.2018.0218 (2019).
77. Donovan, B. M. *et al.* Development and validation of a clinical model for preconception and early pregnancy risk prediction of gestational diabetes mellitus in nulliparous women. *PLoS One* **14**(4), e0215173. https://doi.org/10.1371/journal.pone.0215173 (2019).
78. Punnose, J. *et al.* Glycated haemoglobin in the first trimester: A predictor of gestational diabetes mellitus in pregnant Asian Indian women. *Diabetes Res. Clin. Pract.* **159**, 107953. https://doi.org/10.1016/j.diabres.2019.107953 (2020).
79. Xiong, Y. *et al.* Prediction of gestational diabetes mellitus in the first 19 weeks of pregnancy using machine learning techniques. *J. Matern. Fetal Neonatal Med.* **35**, 1–7. https://doi.org/10.1080/14767058.2020.1786517 (2020).
80. Feng, P., Wang, G., Yu, Q., Zhu, W. & Zhong, C. First-trimester blood urea nitrogen and risk of gestational diabetes mellitus. *J. Cell Mol. Med.* **24**(4), 2416–2422. https://doi.org/10.1111/jcmm.14924 (2020).
81. Zhang, Y. Z. *et al.* A mid-pregnancy risk prediction model for gestational diabetes mellitus based on the maternal status in combination with ultrasound and serological findings. *Exp. Ther. Med.* **20**(1), 293–300. https://doi.org/10.3892/etm.2020.8690 (2020).
82. Maternal and infant research center of the birth study group. CIMIGEN. https://www.cimigen.org.mx/home (2022).

Acknowledgements

Author contributions

H.G.R., M.J.R.B., J.O.M., A.M., L.A.M.J., J.L.G., R.S.M., R.M.R., V.G.H., L.M.J., E.R.M., and R.T.C. conceptualized the study. H.G.R., M.J.R.B., J.O.M., A.M., L.A.M.J., J.L.G., R.S.M., R.M.R., L.M.J., E.R.M., and R.T.C. developed the study methodology. A.M. and R.S.M. were involved in software development. H.G.R., M.J.R.B., J.O.M., A.M., L.A.M.J., J.L.G., R.S.M., R.M.R., L.M.J., and R.T.C. validated the study data. H.G.R., M.J.R.B., J.O.M., A.M., and L.M.J. participated in the formal analysis of the study data. H.G.R., M.J.R.B., J.O.M., A.M., L.A.M.J., J.L.G., V.G.H., L.M.J., and R.T.C. performed study investigations. H.G.R., M.J.R.B., J.O.M., A.M., L.A.M.J., J.L.G., R.S.M., R.M.R., L.M.J., L.M.I.C., I.X.R.C., D.A.D.M., F.J.M.V., E.O.V.A., A.E.B.H., D.R.G., M.C.C.R., L.H.G., and O.R.I. provided resources. M.J.R.B., J.O.M., A.M., and L.M.J. were involved in data curation and visualization. L.A.M.J., J.L.G., R.S.M., R.M.R., V.G.H., L.M.I.C., I.X.R.C., D.A.D.M., F.J.M.V., E.O.V.A., A.E.B.H., D.R.G., M.C.C.R., L.H.G., O.R.I., E.R.M., and R.T.C. supervised the study. A.M., L.A.M.J., J.L.G., R.S.M., R.M.R., V.G.H., and R.T.C. were involved in project administration. H.G.R., L.A.M.J., J.L.G., R.S.M., R.M.R., and R.T.C. were involved in funding acquisition. All authors were involved in the writing of the original draft and the review and editing of this manuscript. All authors confirm that they had full access to all the data in this study and accept responsibility for this publication.

Funding

Competing interests

H.G.R., M.J.R.B., J.O.M., A.M., L.A.M.J., J.L.G., R.S.M., R.M.R., V.G.H., and L.M.J. are employees of The Carlos Slim Foundation. L.M.I.C., I.X.R.C., D.A.D.M., F.J.M.V., E.O.V.A., A.E.B.H., D.R.G., M.C.C.R., L.H.G., O.R.I., E.R.M., and R.T.C. have no conflicts of interest to declare.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-023-34126-7.

Correspondence and requests for materials should be addressed to L.A.M.-J.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.