Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 5, No.3, May : 2025

A SURVEY ON PREDICTION OF GESTATIONAL DIABETES USING MACHINE LEARNING

Miss. Vrushali Bhavarilal Shukla, Student, Department Of Computer Science and Engineering, Sipna College Of Engineering and Technology, Amravati, Maharashtra, India.
Dr. H. R. Vyawahare, Professor, Department Of Computer Science and Engineering, Sipna College Of Engineering and Technology, Amravati, Maharashtra, India.

ABSTRACT:
Gestational diabetes mellitus (GDM) is a type of diabetes that occurs during pregnancy and can lead to significant maternal and fetal health complications if left undiagnosed or untreated. Early prediction and intervention are critical to managing the condition effectively. In this study, machine learning (ML) techniques are applied to predict the risk of GDM in pregnant women based on various clinical and demographic factors. Using a dataset consisting of patient information such as age, body mass index (BMI), blood pressure, glucose levels, and previous medical history, various ML models, including logistic regression, decision trees, random forests, and support vector machines, are trained and evaluated for their predictive performance. The models' accuracy, sensitivity, specificity, and area under the receiver operating characteristic (ROC) curve (AUC) are analysed to determine the most effective approach for GDM prediction. Results demonstrate that machine learning algorithms, particularly random forests and support vector machines, provide high accuracy in predicting gestational diabetes, suggesting their potential for clinical use in early screening and decision-making. The findings highlight the value of integrating ML-based prediction tools into routine prenatal care to enhance GDM detection and improve maternal-fetal health outcomes.
Keywords — Gestational Diabetes, Machine Learning, Predictive Modelling, Classification, Healthcare.

INTRODUCTION:
Gestational diabetes mellitus (GDM) is a significant health concern affecting pregnant women, with potential risks including pre-eclampsia, fetal macrosomia, and long-term development of type 2 diabetes in the mother. According to the World Health Organization, the prevalence of GDM has been rising globally. Early detection of GDM allows for timely intervention, such as dietary modifications and insulin therapy, reducing adverse outcomes.
Traditional methods of GDM screening include oral glucose tolerance tests (OGTT), but these are often limited by high costs, invasiveness, and delayed results. Machine learning (ML) offers an alternative for early diagnosis by leveraging patient data for predictive modelling. Gestational diabetes (GDM) can be broadly defined as glucose intolerance during pregnancy that affects women without previous diagnosis of diabetes or unknown state. The incidence is about 7% worldwide and this rate has been growing during the last decades and is estimated to increase in the future. The most important risk factors are maternal overweight and obesity, age greater than or equal to 35 years at delivery, hypertension, metabolic syndrome, nonwhite ethnicity, family history of diabetes mellitus, prior unexplained stillbirth, prior infant with congenital anomaly (if not screened during that pregnancy), prior macrocosmic infant, history of gestational diabetes, chronic use of steroids, glycosuria, and known impaired glucose metabolism. The importance of GDM is linked to the consequences of pregnancy and also after pregnancy to both mother and newborn. Hyperglycaemia in the mother causes abnormal metabolism while in the fetus it causes hyperinsulinemia and its consequences, and incidence of complications is inversely proportional to glucose control. Macrosomia, polyhydramnios, operative delivery, shoulder dystocia, birth injury, perinatal mortality, hypertensive disorders and preeclampsia, congenital malformations (OR: 1.2–1.4), and risk of caesarean delivery are higher in women with GDM; in the long term, women with GDM have a higher risk of developing type 2 diabetes mellitus

and cardiovascular diseases; long-term sequelae for offspring are obesity and metabolic syndrome. Approximately 50% of women identified a shaving GDM will develop frank diabetes within 10 years. To prevent or decrease the risk of GDM, weight loss before pregnancy and cardiovascular exercise could be useful. In fact, aerobic exercise for 35–90 minutes 3-4 times per week during pregnancy is associated with a significantly higher incidence of vaginal delivery and a significantly lower incidence of caesarean delivery, with a significantly lower incidence of gestational diabetes mellitus and hypertensive disorders. Prompt diagnosis and management are important to reduce worse pregnancy outcomes. Nonetheless, screening, management, and follow-up of GDM are controversial on international organizations recommendations.

Overview of Gestational Diabetes:

GDM is characterized by glucose intolerance that occurs during pregnancy. It typically manifests between the 24th and 28th weeks of gestation and is diagnosed through blood glucose testing after oral glucose intake. Factors such as maternal age, family history, obesity, and ethnicity influence the likelihood of developing GDM.

Machine Learning in Healthcare:

Machine learning has made significant strides in healthcare, particularly in disease prediction and diagnostic applications. The ability of ML models to learn complex patterns from large datasets has facilitated improvements in early detection, risk assessment, and personalized treatment strategies.

LITERATURE REVIEW:

Previous Work on GDM Prediction:

Machine learning methods for GDM prediction have been investigated in a number of papers. Numerous classifiers, including as decision trees, support vector machines (SVM), random forests, and neural networks, have been used in these investigations. The Gestational Diabetes Dataset from the UCI Machine Learning Repository and the Pima Indians Diabetes Database are often used datasets. The models are frequently assessed using performance criteria such area under the receiver operating characteristic (ROC) curve, sensitivity, specificity, and accuracy.

Claudia et al. [1] We found very little evidence about whether treating women who meet the IADPSG criteria (One-Step test) for GDM but not by other less stringent criteria has an impact on unfavourable pregnancy outcomes when compared to no treatment, despite ongoing debate about whether the One Step or Two-Step test should be used for GDM screening. Furthermore, the study group with milder condition was not treated for GDM (positive for IADPSG criteria, but negative for less strict criteria) in any of the included trials. Additionally, we discovered that the literature used a wide range of criteria (IADPSG, WHO, NICE, CDA, and C&C) for GDM screening.

Therefore, it is not unexpected that certain societies, like ACOG, continue to suggest the Two-Step technique for screening, while others, like IADPSG, WHO, and FIGO, recommend the One Step approach (thinking that identifying women with milder GDM would have benefits for them and their newborns). This question could only be addressed by well-planned RCTs that included large populations and compared the One-Step and Two-Step approaches. Regarding GDM management, there are a lot of unanswered questions. After conducting a thorough literature analysis, we discovered various standards for GDM screening, GDM monitoring, and the initiation of pharmaceutical treatment. The goal is to arrive at widely accepted and agreed-upon recommendations to enhance healthcare, lower expenses, and lessen negative consequences for women with GDM and their unborn children.

Pouya et al. [2] The IDF Diabetes Atlas emphasises the need for immediate action to address diabetes prevention, early detection, and care, as well as the disease's increasing worldwide burden. The significance of public health programs aimed at risk factors like obesity, poor food, and inactivity is highlighted by the sharp rise in prevalence. Around 50% of diabetics worldwide went undiagnosed in

2019, highlighting a major obstacle to early detection and prevention. Thomas, R.L. et al. [3] An Overview of Research on the Global Prevalence of Diabetes-Related Retinopathy Using Retinal Photography examines a number of studies conducted between 2015 and 2018 that evaluated the prevalence of diabetic retinopathy (DR) worldwide within the given time frame using retinal photography. The review sheds light on the extensive effects of DR on diabetics and emphasises how retinal imaging has emerged as a crucial diagnostic and monitoring tool. Based on variables like geographic location, population demographics, and healthcare availability, the research under review demonstrate differing incidence rates across a range of locales. The analysis highlights the need for better screening, early identification, and intervention options to avoid vision loss in diabetic individuals worldwide, as well as the growing global burden of DR as diabetes prevalence rises.Increasing Prevalence of Gestational Diabetes Mellitus [4] Assiamira Ferrara talks on the growing prevalence of gestational diabetes (GDM), a disorder in which high blood sugar levels occur during pregnancy, worldwide. It draws attention to the causes of the rise, including shifting lifestyle choices, an increase in obesity, and ageing mothers. The analysis highlights the health concerns that mothers and babies face, such as delivery difficulties and the chance of type 2 diabetes in later life. In order to control and lessen the effects of GDM on the health of mothers and children, the article advocates for better screening, early identification, and preventive measures.Dahanayaka, N. J., and others [5] The limits of employing a risk factor-based screening technique for gestational diabetes mellitus (GDM) are criticised in Inadequacy of the Risk Factor-Based Approach to Detect Gestational Diabetes Mellitus. It makes the case that focussing only on variables like age, obesity, and family history frequently misses many women who are at risk, resulting in under diagnosis and lost chances for early intervention. In order to ensure prompt detection and improved outcomes for both mothers and newborns, the review emphasises that GDM can impact women who do not have typical risk factors and promotes more widespread screening techniques, such as universal glucose testing. The diagnosis and classification of diabetes mellitus, encompassing Type 1, Type 2, gestational diabetes, and other less prevalent variants, is summarised in A. Sumathi et al. [6] Diagnosis and Classification of Diabetes Mellitus. It describes the importance of the main diagnostic tests in identifying diabetes, including haemoglobin A1c (HbA1c), oral glucose tolerance test (OGTT), and fasting plasma glucose (FPG). Updated diagnostic standards from groups like the World Health Organisation (WHO) and the American Diabetes Association (ADA) are also included in the article. It also emphasises how crucial precise classification is for directing therapeutic approaches, controlling side effects, and stopping the progression of the disease.Variable blood glucose levels during pregnancy are a hallmark of gestational diabetes mellitus (GDM) [7]. According to one recent study, GDM is rapidly spreading over the world and affects Chinese pregnant women. GDM mothers are at increased risk for preeclampsia, placental malfunction, metabolic disruption, and caesarean birth. Reduced foetal growth and increased chances of shoulder dystocia, preterm birth, macrosomia, and birth trauma are the outcomes of hyperglycaemia and placental malfunction [8]. Obesity, type 2 diabetes, and heart attacks are among the postpartum consequences that a mother with GDM and her newborn child face. In order to lower the incidence of GDM and low unfavourable pregnancy, early diagnosis and prevention are crucial [9]. However, the majority of GDM cases are identified by the Oral Glucose Tolerance Test (OGTT) between weeks 24 and 28 of pregnancy. Given the pre-existing foetal and placental growth, this is an appropriate window for intervention. Early in pregnancy, the OGTT diagnosis approach was advised by the previous study [10]. However, because GDM typically appears during mid-to-late delivery, it is costly and useless in the majority of cases. Therefore, early in pregnancy, a basic model should be presented with the aid of conventional medical knowledge. Measuring the risks of GDM and identifying high-risk mothers who need early treatment, monitoring, and medication should be made easier using this methodology. In this manner, universal OGTTs among women at low risk can be decreased. Classical regression analysis is used in conjunction with newly discovered detection techniques to forecast GDM. However,

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 5, No.3, May : 2025

by "learning" from the data, machine learning (ML), a data analytics technique, builds the model for forecasting outcomes. This method has been highlighted as a capable substitute for regression analysis. Additionally, ML can outperform classical regression, conceivably due to its ability to capture complex interactions between predictive characteristics and nonlinearities [11]. Even though there have been the most studies in this field, very few have used machine learning (ML) to predict GDM, and no models have been compared to logistic regressions (LR). Xiong et al. [12] made the decision to use Support Vector Machine (SVM) and light Gradient Boosting Machine (light GBM) to create a risk prediction mechanism for the first 19 weeks with high-potential GDM predictors. A straightforward strategy for identifying GDM in an early stage of pregnancy utilising biochemical markers and machine learning was provided by Zheng et al. [13]. In order to construct an artificial intelligence (AI)-based application, Shen et al.'s study [14] noted that the evaluation of the best AI approach in GDM prediction requires the fewest clinical equipment and trainees. The PIMA dataset is used to predict GDM using various machine learning techniques in the literature [15]. Therefore, applied performance indicators were used to assess the accuracy of ML models. In the administration of the diabetes PIMA data set, the confusion matrix, Receiver Operating Characteristic (ROC), and AUC measurements help to understand the importance of the machine learning technique. A statistical approach for estimating GDM with Microsoft Azure AI services was presented by Srivastava et al. [16]. It is known as ML Studio and uses the drag-and-drop method to get best performance. Additionally, this study forecasted the occurrence of GDM based on characteristics implicated in early stages of pregnancy using a classification algorithm. To create the prediction schemes, the study took into account five traditional machine learning techniques as well as the Cost-Sensitive Hybrid Model (CSHM). Temporary Electronic Health Records (EHRs) have been used by the writers of the literature [17] to study the future dangers of GDM. A small amount of data must be recorded and gathered in order to assemble the dataset once the data cleaning process is finished. The Radial Basis Function Network (RBF Network), an artificial neural network (ANN) technique, was created by the authors of the literature [18], and performance validation and comparison analysis were carried out. This technique was used to find potential cases of GDM, which can pose a number of dangers to both the foetus and the pregnant woman. Ye et al. [19] used a variety of ML and traditional LR techniques to train their parameters. Three different classifiers were used by Du et al. [20] to forecast future GDM risk. This study's detection accuracy aids the clinician in making the best choices possible, which makes it easier to prevent the condition. According to the study, the DenseNet model has the highest degree of flexibility in detecting GDM. An ensemble of ML-based GDM prediction and classification models is presented in the current study. Three processes—preprocessing, classification, and ensemble voting—are included in the model that is being described. The researcher employed four machine learning models for classification: Random Forest (RF), Support Vector Machine (SVM), k-Nearest Neighbour (KNN), and Logistic Regression (LR). Additionally, a voting classifier was employed in conjunction with RF, LR, and SVM classifiers to complete the final classification. The findings of the analysis obtained from the offered model were compared with conventional approaches in order to verify the effectiveness of the suggested approach. Additionally, a large number of experiments were carried out on many topics.

METHODOLOGY:

Data Collection

This study utilizes publicly available datasets such as the **Pima Indians Diabetes Dataset** and the **Gestational Diabetes Database** from UCI. The datasets include features such as age, BMI, blood pressure, glucose levels, insulin, and family history, which are commonly associated with gestational diabetes.

PREPROCESSING:
Data preprocessing steps include handling missing values, normalization, and feature selection. For this study, missing values are imputed using the mean for numerical features and the mode for categorical features. Features are normalized to a standard scale to ensure fair model performance.

MODEL SELECTION:
We test a number of machine learning models:
- **Logistic Regression (LR)**: A statistical method for binary classification that is simple yet effective.
- **Support Vector Machine (SVM)**: A powerful classifier that works well with high-dimensional data.
- **Random Forest (RF)**: An ensemble learning method that combines multiple decision trees to improve accuracy and prevent overfitting.
- **K-Nearest Neighbors (KNN)**: A non-parametric method that classifies based on proximity to labeled examples.
- **Artificial Neural Networks (ANN)**: A deep learning approach that learns complex patterns in data through multiple hidden layers.

EVALUATION METRICS:
The models' performance is assessed using:
- **Accuracy**: The overall proportion of correctly classified instances.
- **Sensitivity**: The proportion of true positives correctly identified.
- **Specificity**: The proportion of true negatives correctly identified.
- **Area Under the ROC Curve (AUC)**: A metric for evaluating the overall performance of a binary classification model.

CONCLUSION:
Machine learning techniques offer promising solutions for predicting gestational diabetes, providing faster and potentially more accurate assessments compared to traditional diagnostic methods. Among the models tested, Random Forest demonstrated the best overall performance. Future work should focus on improving the interpretability of machine learning models and exploring deep learning methods for handling larger, more complex datasets. Additionally, the integration of these models into clinical practice could lead to early and personalized intervention strategies for pregnant women at risk of GDM.

REFERENCES:
1. Caissutti C, Berghella V. Scientific Evidence for Different Options for GDM Screening and Management: Controversies and Review of the Literature. Biomed Res Int. 2017;2017:2746471. doi: 10.1155/2017/2746471. Epub 2017 Apr 10. PMID: 28497042; PMCID: PMC5402236.
2. Saeedi P, Petersohn I, Salpea P, Malanda B, Karuranga S, Unwin N, Colagiuri S, Guariguata L, Motala AA, Ogurtsova K, Shaw JE, Bright D, Williams R; IDF Diabetes Atlas Committee. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9th edition. Diabetes Res Clin Pract. 2019 Nov;157:107843. doi: 10.1016/j.diabres.2019.107843. Epub 2019 Sep 10. PMID: 31518657.
3. Thomas RL, Halim S, Gurudas S, Sivaprasad S, Owens DR. IDF Diabetes Atlas: A review of studies utilising retinal photography on the global prevalence of diabetes related retinopathy between 2015 and 2018. Diabetes Res Clin Pract. 2019 Nov;157:107840. doi: 10.1016/j.diabres.2019.107840. Epub 2019 Nov 14. PMID: 31733978.

4. Assiamira Ferrara, “Increasing Prevalence of Gestational Diabetes Mellitus: A public health perspective”, Diabetes Care 2007;30(Supplement_2):S141–S146 https://doi.org/10.2337/dc07-s206, PubMed: 17596462.
5. Dahanayaka NJ, Agampodi SB, Ranasinghe OR, Jayaweera PM, Wickramasinghe WA, Adhikari AN, Chathurani HK, Dissanayaka UT. Inadequacy of the risk factor based approach to detect gestational diabetes mellitus. Ceylon Med J. 2012 Mar;57(1):5-9. doi: 10.4038/cmj.v57i1.4193. PMID: 22453704.
6. Sumathi*,” Ensemble Classifier Technique to Predict Gestational Diabetes Mellitus (GDM)” Computer Systems Science and Engineering 2022, 40(1),325. https://doi.org/10.32604/csse.2022.017484 **Received** 31 January 2021; **Accepted** 23 April 2021; **Issue published** 26 August 2021
7. X. Mao, X. Chen, C. Chen, H. Zhang and K. P. Law, “Metabolomics in gestational diabetes,” Clinica Chimica Acta, vol. 475, pp. 116–127, 2017.
8. M. L. Geurtsen, E. E. L. V. Soest, E. Voerman, E. A. P. Steegers, V. W. V. Jaddoe et al., “High maternal early pregnancy blood glucose levels are associated with altered fetal growth and increased risk of adverse birth outcomes,” Diabetologia, vol. 62, no. 10, pp. 1880–1890, 2019.
9. C.E.Powe, “Early pregnancy biochemical predictors of gestational diabetes mellitus,” Current Diabetes Reports, vol. 17, no. 2, pp. 1–10, 2017.
10. S. Shinar and H. Berger, “Early diabetes screening in pregnancy,” International Journal of Gynecology & Obstetrics, vol. 142, no. 1, pp. 1–8, 2018.
11. D.D.Miller and E. W. Brown, “Artificial intelligence in medical practice: the question to the answer?,” American Journal of Medicine, vol. 131, no. 2, pp. 129–133, 2018.
12. Y.Xiong, L. Lin, Y. Chen, S. Salerno, Y. Li et al., “Prediction of gestational diabetes mellitus in the first 19 weeks of pregnancy using machine learning techniques,” Journal of Maternal-Fetal & Neonatal Medicine, vol. 33, no. 1, pp. 1–8, 2020.
13. T. Zheng, W. Ye, X. Wang, X. Li, J. Zhang et al., “A simple model to predict risk of gestational diabetes mellitus from 8 to 20weeks ofgestation in chinese women,” BMC Pregnancy and Childbirth, vol. 19, no. 1, pp. 1–10, 2019.
14. J. Shen, J. Chen, Z. Zheng, J. Zheng, Z. Liu et al., “An innovative artificial intelligence-based app for the diagnosis of gestational diabetes mellitus (GDM-AI): development study,” Journal of Medical Internet Research, vol. 22, no. 9, pp. e21573, 2020.
15. I.Gnanadass, “Prediction of gestational diabetes by machine learning algorithms,” IEEE Potentials, vol. 39, no. 6, pp. 32–37, pp. 1–1, 2020.
16. Y. Srivastava, P. Khanna and S. Kumar, “February. estimation of gestational diabetes mellitus using azure AI services,” in 2019 Amity Int. Conf. on Artificial Intelligence (AICAI) IEEE, Dubai, United Arab Emirates, pp. 323–326, 2019.
17. H. Qiu, H. Y. Yu, L. Y. Wang, Q. Yao, S. N. Wu et al., “Electronic health record driven prediction for gestational diabetes mellitus in early pregnancy,” Scientific Reports, vol. 7, no. 1, pp. 1–13, 2017.
18. M. W. Moreira, J. J. Rodrigues, N. Kumar, J. A. Muhtadi and V. Korotaev, “Evolutionary radial basis function network for gestational diabetes data analytics,” Journal of Computational Science, vol. 27, pp. 410–417, 2018.
19. Y. Ye, Y. Xiong, Q. Zhou, J. Wu, X. Li et al., “Comparison of machine learning methods and conventional logistic regressions for predicting gestational diabetes using routine clinical data: a retrospective cohort study,” Journal of Diabetes Research, vol. 2020, pp. 1–10, 2020.
20. F. Du, W. Zhong, W. Wu, D. Peng, T. Xu et al., “Prediction of pregnancy diabetes based on machine learning,” in BIBE 2019; The Third Int. Conf. on Biological Information and Biomedical Engineering, Hangzhou, China, pp. 1–6, 2019.