Gestational Diabetes Finder Using Machine Learning

1 SASIKUMAR A N, 2AVINASH S

1Assistant Professor, 2,Undergraduate Student

Department of Computer Science and Engineering,

Panimalar Engineering College, Chennai, India

1ansasikumar@gmail.com,2avinash866749@gmail.com,

*Abstract*— **Gestational diabetes is a type of diabetes that occurs only during pregnancy. Gestational diabetes can cause health problems in both mother and baby. Managing your diabetes can help protect you and your baby. Gestational diabetes often has no symptoms, or they may be mild, such as being thirstier than normal or having to urinate more often. Gestational diabetes is sometimes related to the hormonal changes of pregnancy that make your body less able to use insulin. Genes and extra weight may also play a role.Your doctor will test you for gestational diabetes between 24 and 28 weeks of pregnancy. Tests include the glucose challenge test and the oral glucose tolerance test (OGTT). If the results of the glucose challenge test show high blood glucose, you will return for an OGTT test to confirm the diagnosis of gestational diabetes.Using machine learning algorithms the patient data is obtained and read further processed using suitable algorithms, Random forest models therefore calculating the accuracy levels and comparing it and then used for prediction.By developing a system which shows perfect accuracy level and could save the life of mother and fetus from disease. Predicting the early stage would be easier to treat the patient with medications and control with precautionaries, further controlling complications.**

*Index Terms*—Gestational Diabetes , Machine learning, Random Forest Model , OGTT.

*\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*

# **Introduction**

Gestational diabetes means high blood sugar levels during pregnancy. In many women, it is a temporary condition that goes away after birth. However, if the women are at risk of developing gestational diabetes, they should work with the doctor to manage their blood sugar levels throughout pregnancy. They can look for an easy high blood sugar during pregnancy meal plan and exercise regularly to maintain healthy blood sugar levels. Maintaining blood sugar levels during pregnancy keeps your pregnancy and baby away from several health-related complications.

Giving birth is one of the most beautiful experiences in the life of a woman. However, this experience can be bitter if a woman has diabetes or develops diabetes during pregnancy. A lot of women with diabetes may have to face severe health complications during and post-delivery. Diabetes during pregnancy affects the mother and also increases the risk for the baby in the womb. Therefore, it is significant to keep the blood sugar levels under control throughout pregnancy to have a healthy pregnancy and healthy babies.By developing a system which shows perfect accuracy level and could save the life of mother and fetus from disease. Predicting the early stage would be easier to treat the patient with medications and control with precautionaries, further controlling complications.

Managing gestational diabetes includes following a healthy eating plan and being physically active. If your eating plan and physical activity aren’t enough to keep your blood glucose in your target range, you may need insulin.

You can lower your chance of getting gestational diabetes by losing extra weight before you get pregnant if you are overweight. Being physically active before and during pregnancy also may help prevent gestational diabetes.

Managing gestational diabetes includes following a healthy eating plan and being physically active. If your eating plan and physical activity aren’t enough to keep your blood glucose in your target range, you may need insulin.

Risk factors for GDM

* Several risk factors are associated with the development of GDM.
* The most common risk factors are; obesity, older maternal age, past history of GDM, strong family history of diabetes, member of an ethnic group with a high prevalence of T2DM, polycystic ovary syndrome, and persistent glucosuria
* A history of delivering big baby (birth weight ≥4000 g), history of recurrent abortions, and history of unexplained stillbirths, and history of essential hypertension, or pregnancy-related hypertension are other risk factors for GDM.

**Risks of GDM**

* Women with GDM have an increased incidence of hypertensive disorders during pregnancy, including gestational hypertension, pre-eclampsia, and eclampsia.
* There is an increase risk of polyhydramnios that may increase the risk of preterm labor.
* Excessive fetal growth remains an important perinatal concern in GDM.
* Consequences of excessive fetal growth include birth trauma, maternal morbidity from cesarean deliveries, shoulder dystocia, and neonatal hypoglycemia.
* Other neonatal morbidities that potentially occur more frequently in infants of women with GDM include hyperbilirubinemia, hypocalcemia, erythema, and respiratory distress syndrome.
* Long-term complications of GDM include diabetes and cardiovascular disease in the mothers, and obesity and diabetes in the offspring

# **Literature Survey**

The following authors and their papers are taken into account in the development of this research paper.

1. **Du, Y., Rafferty, A.R., McAuliffe, F.M. *et al.* An explainable machine learning-based clinical decision support system for prediction of gestational diabetes mellitus. *Sci Rep* 12, 1170 (2022). https://doi.org/10.1038/s41598-022-05112-2[1]** In this paper, We have developed an explainable machine learning-based clinical decision support system (CDSS) to identify at-risk women in need of targeted pregnancy intervention. Maternal characteristics and blood biomarkers at baseline from the PEARS study were used. After appropriate data preparation, synthetic minority oversampling technique and feature selection, five machine learning algorithms were applied with five-fold cross-validated grid search optimising the balanced accuracy. Our models were explained with Shapley additive explanations to increase the trustworthiness and acceptability of the system. We developed multiple models for different use cases: theoretical (AUC-PR 0.485, AUC-ROC 0.792), GDM screening during a normal antenatal visit (AUC-PR 0.208, AUC-ROC 0.659), and remote GDM risk assessment (AUC-PR 0.199, AUC-ROC 0.656). Our models have been implemented as a web server that is publicly available for academic use. Our explainable CDSS demonstrates the potential to assist clinicians in screening at risk patients who may benefit from early pregnancy GDM prevention strategies
2. **Zhang Z, Yang L, Han W, Wu Y, Zhang L, Gao C, Jiang K, Liu Y, Wu H Machine Learning Prediction Models for Gestational Diabetes Mellitus: Meta-analysisJ Med Internet Res 2022;24(3):e26634doi: [10.2196/26634](https://doi.org/10.2196/26634)PMID: [35294369](https://www.ncbi.nlm.nih.gov/pubmed/35294369) [2]** In this study ,we have perform a meta analysis and comparison of published prognostic models for predicting the risk of GDM and identify predictors applicable to the models.Four reliable electronic databases were searched for studies that developed ML prediction models for GDM in the general population instead of among high-risk groups only. The novel Prediction Model Risk of Bias Assessment Tool (PROBAST) was used to assess the risk of bias of the ML models. The Meta-DiSc software program (version 1.4) was used to perform the meta-analysis and determination of heterogeneity. To limit the influence of heterogeneity, we also performed sensitivity analyses, a meta-regression, and subgroup analysis.
3. **Hang Lai, Huaxiong Huang, Karim Keshavjee, Aziz Guergachi & Xin Gao. Predictive models for diabetes mellitus using machine learning techniques.**The objective of this study was to build an effective predictive model with high sensitivity and selectivity to better identify Canadian patients at risk of having Diabetes Mellitus based on patient demographic data and the laboratory results during their visits to medical facilities.Using the most recent records of 13,309 Canadian patients aged between 18 and 90 years, along with their laboratory information (age, sex, fasting blood glucose, body mass index, high-density lipoprotein, triglycerides, blood pressure, and low-density lipoprotein), we built predictive models using Logistic Regression and Gradient Boosting Machine (GBM) techniques. The area under the receiver operating characteristic curve (AROC) was used to evaluate the discriminatory capability of these models. We used the adjusted threshold method and the class weight method to improve sensitivity – the proportion of Diabetes Mellitus patients correctly predicted by the model. We also compared these models to other learning machine techniques such as Decision Tree and Random Forest**.**
4. **Zou, Yan; Gong, Xue; Miao, Puyang; Liu, Yan (2020). [IEEE 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC) - Chongqing, China (2020.6.12-2020.6.14)] 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC) - Using TensorFlow to Establish multivariable linear regression model to Predict Gestational Diabetes\*. , (), 1695–1698. doi:10.1109/itnec48623.2020.9084664** In this research, we collected data of obstetrics and gynaecology clinic in Chifeng Hongshan District Maternal and child health care centre. After data cleaning and pretreatment , a total of 419 pregnancy data sets were collected . The body mass index(BMI), the pre-pregnancy BMI, the maternal age and the family history of diabetes and the fetal abdominal circumference on the day of pregnancy were selected as independent variables , and a multivariable linear regression model was used to learn . In the predictive model of diabetes, the least squares method and stochastic gradient descent algorithm were used to optimize the training and predictive model was evaluated.
5. **Zou Q, Qu K, Luo Y, Yin D, Ju Y and Tang H (2018) Predicting Diabetes Mellitus With Machine Learning Techniques. *Front. Genet.* 9:515. doi: 10.3389/fgenLearning Framework for Robust Obstacle Detection, Recognition, and Tracking, Vinh Dinh Nguyen, Hau Van Nguyen, Dinh Thi Tran, Sang Jun Lee, and Jae Wook Jeon, Member, IEEE, [3]** In this study, we used decision tree, random forest and neural network to predict diabetes mellitus. The dataset is the hospital physical examination data in Luzhou, China. It contains 14 attributes. In this study, five-fold cross validation was used to examine the models. In order to verity the universal applicability of the methods, we chose some methods that have the better performance to conduct independent test experiments. We randomly selected 68994 healthy people and diabetic patients’ data, respectively as training set. Due to the data unbalance, we randomly extracted 5 times data. And the result is the average of these five experiments. In this study, we used principal component analysis (PCA) and minimum redundancy maximum relevance (mRMR) to reduce the dimensionality. The results showed that prediction with random forest could reach the highest accuracy (ACC = 0.8084) when all the attributes were used**.**

# **Proposed Work**

1. System Architecture

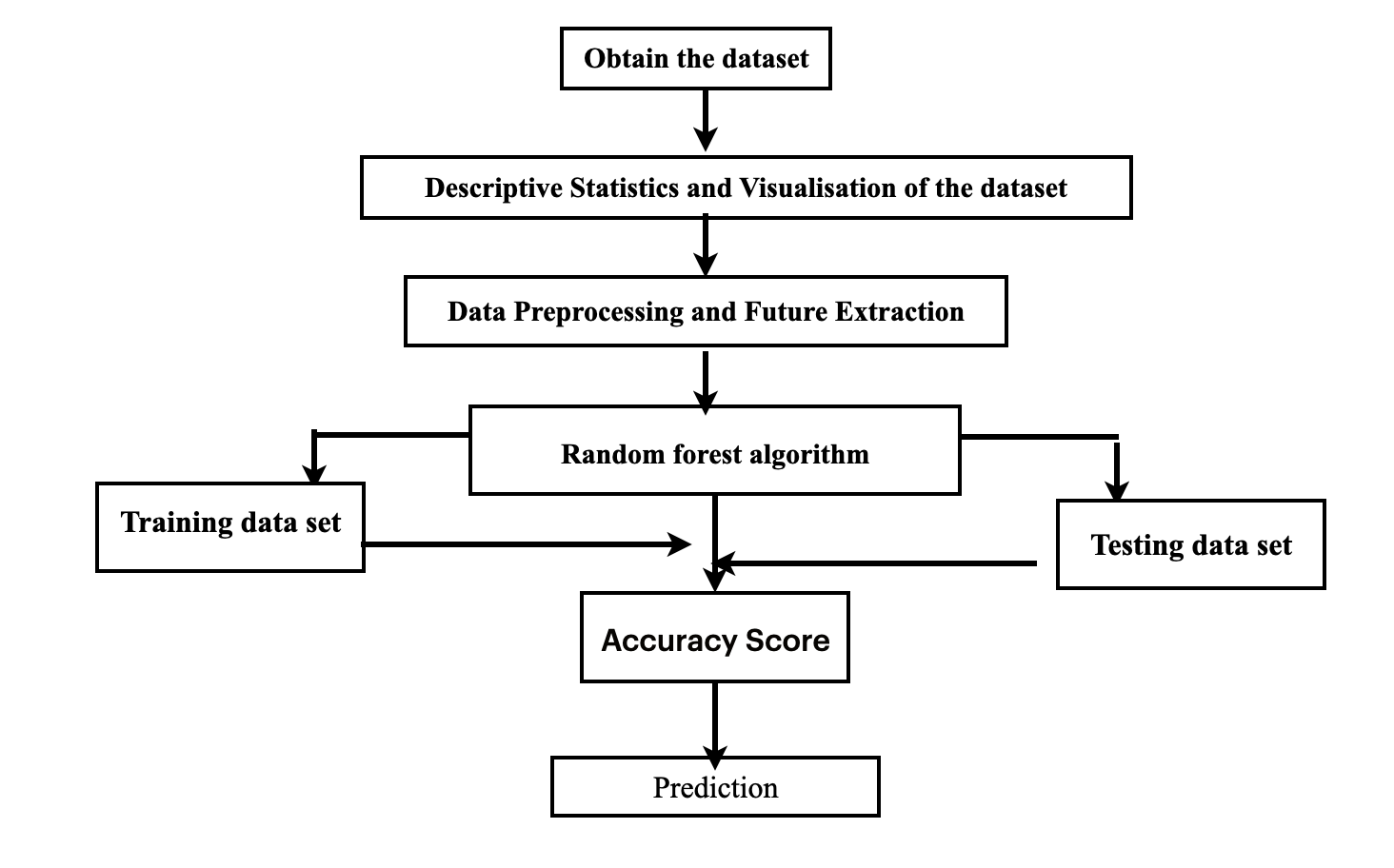


Fig. 3. Architecture Diagram

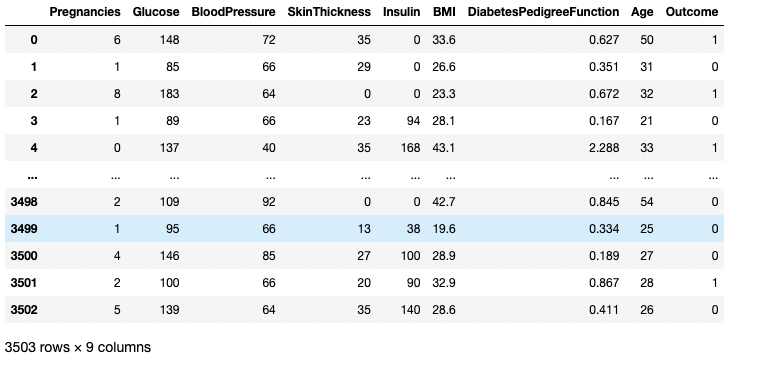
# **Steps involved in identification of gestational diabetes**

In this paper, Gestational Diabetes finder has developed as a Machine Learning application. In this section we shall learn about the classifier used in machine learning to predict the diabetes. We shall also explain our proposed methodology to improve the accuracy . Random Forest algorithm used in this paper. The output is the accuracy metrics of the machine learning models. Then, the model can be used in prediction.

**Data Collection**

A data set is a collection of data. Machine Learning has become the go-to method for solving many challenging real-world problems. It’s definitely by far the best performing method for handling huge amount of data and providing solution. The two diabetes dataset were originated from [www.kaggle.com](http://www.kaggle.com) .The first dataset which was originated from kaggle is (“<https://www.kaggle.com/johndasilva/diabetes>) the dataset was extracted from a hospital in Germany named Frankfurt.

The second data which was originated is PIMA dataset(This dataset is originally from the National Institute of Diabetes and Digestive and Kidney Diseases. The objective of the dataset is to diagnostically predict whether or not a patient has diabetes, based on certain diagnostic measurements included in the dataset. Several constraints were placed on the selection of these instances from a larger database. In particular, all patients here are females at least 21 years old of Pima Indian heritage )which was extracted from kaggle website . By combining the two dataset the accuracy level of the model is hit and prediction are accurate too. These Machine learning models that have been working so well because they need lots of fuel; that fuel is data. The more **labelled data** available, the better our model performs.

****

* The dataset contains of **3504** datapoints with **9** features each**(Pregnancies,Glucose,**

**Blood Pressure, Skin-thickness , Insulin, BMI, DPF, Age ,Outcomes**)

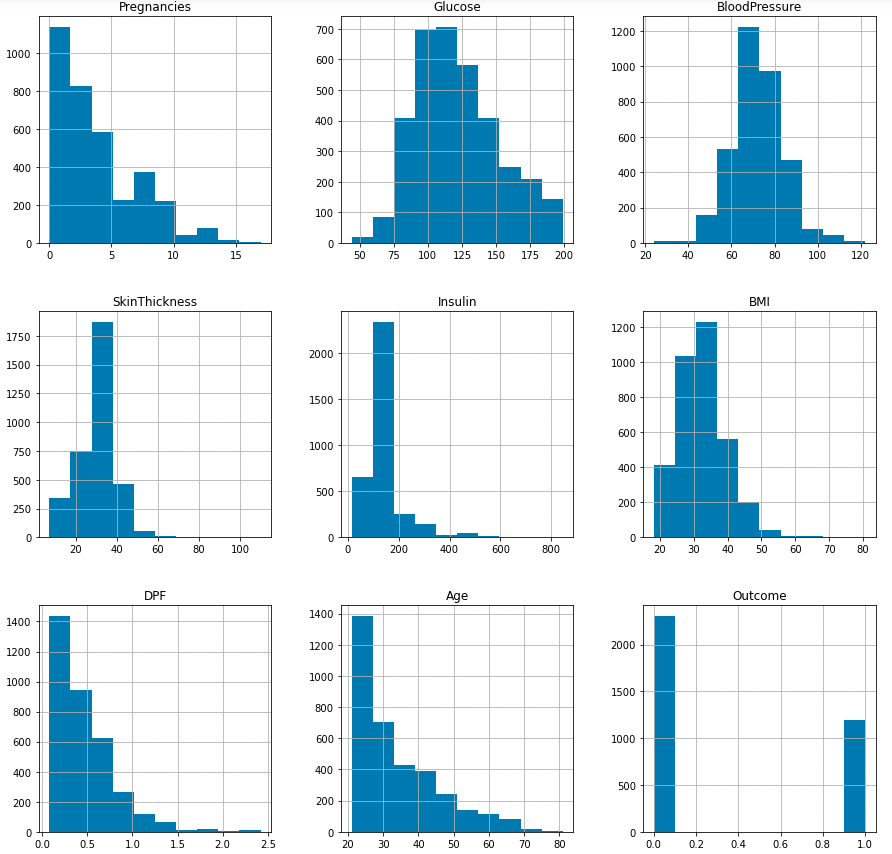
* The outcome “**0**” it denotes the patients doesn’t have diabetes , while the outcome”**1**” denotes the patients have diabetes.

Image

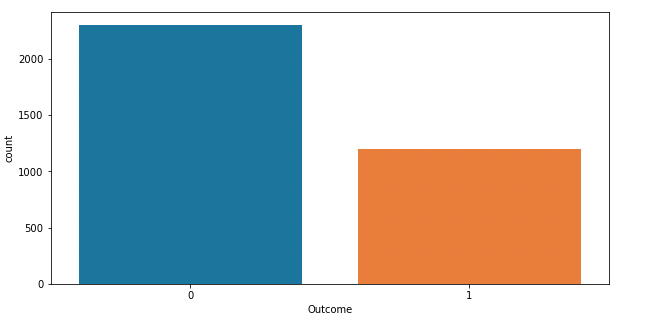
**Proposed Model Diagram**

# **Result**

To begin with building a model , the data should be cleansed and pre processed to perform prediction .So, we can split our project into modules of implementation that is done.

In the dataset, many values are entered as “0” so the accuracy level could decrease and the prediction may be negative . As a result , the value which are replaced with NaN value and then the NaN values are filled using the strategies like mean, median, most\_frequent, constant.

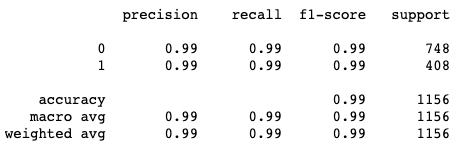
Let's take a look at the plots . It shows how each feature and label is distributed along different ranges, which further confirms the need for scaling . The main thing to be remembered is, whenever you see discrete bars, it basically means that each of these is actually a categorical value. We will need to handle the categorical value with cautious before applying machine learning.

**Bar plot for Outcome class**

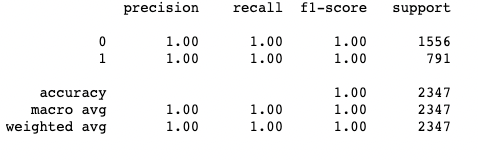
The above graph shows the data that the non -diabetics is almost twice the number of diabetic patients.

**Random forest**

It creates a forest of trees where each tree is formed by a random selection of features from the total features.Image



**Testing Accuracy**



Image

**Training Accuracy**

**VI Conclusion**

one of the important real-world medical problems is the detection of diabetes at it early stage.In this study project has been successfully implemented to predict the gestational diabetes. Experiments are performed on John diabetes database and PIMA dataset .

**VII References**

1. F. Du et al., "Prediction of pregnancy diabetes based on machine learning," BIBE 2019; The Third International Conference on Biological Information and Biomedical Engineering, 2019, pp. 1-6.
2. I. Gnanadass, "Prediction of Gestational Diabetes by Machine Learning Algorithms," in IEEE Potentials, vol. 39, no. 6, pp. 32-37, Nov.-Dec. 2020, doi: 10.1109/MPOT.2020.3015190.
3. V. Ganesh, J. Kolluri and K. V. Kumar, "Diabetes Prediction using Logistic Regression and Feature Normalization," 2021 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES), 2021, pp. 1-6, doi: 10.1109/ICSES52305.2021.9633773.
4. R. S. Shankar, V. V. S. Raju, K. Murthy and D. Ravibabu, "Optimized Model for Predicting Gestational Diabetes using ML Techniques," 2021 5th International Conference on Electronics, Communication and Aerospace Technology (ICECA), 2021, pp. 1623-1629, doi: 10.1109/ICECA52323.2021.9676075.
5. Alfadhli EM. Gestational diabetes mellitus. Saudi Med J. 2015;36(4):399-406. doi:10.15537/smj.2015.4.10307
6. Alfadhli, Eman M. “Gestational diabetes mellitus.” Saudi medical journal vol. 36,4 (2015): 399-406. doi:10.15537/smj.2015.4.10307
7. Du, Y., Rafferty, A.R., McAuliffe, F.M. et al. An explainable machine learning-based clinical decision support system for prediction of gestational diabetes mellitus. Sci Rep 12, 1170 (2022). <https://doi.org/10.1038/s41598-022-05112-2>
8. Zhang Z, Yang L, Han W, Wu Y, Zhang L, Gao C, Jiang K, Liu Y, Wu H Machine Learning Prediction Models for Gestational Diabetes Mellitus: Meta-analysis JMed Internet Res 2022;24(3):e26634doi: [10.2196/26634](https://doi.org/10.2196/26634)
9. Jyotismita Chaki, S. Thillai Ganesh, S.K Cidham, S. Ananda Theertan, Machine learning and artificial intelligence based Diabetes Mellitus detection and self-management: A systematic review <https://doi.org/10.1016/j.jksuci.2020.06.013>
10. Aljumah, A.A., Ahamad, M.G., Siddiqui, M.K., 2013. Application of data mining: Diabetes health care in young and old patients. Journal of King Saud University - Computer and Information Sciences 25, 127–136.doi:10.1016/j.jksuci.2012.10.003.
11. Arora, R., Suman, 2012. Comparative Analysis of Classification Algorithms on Different Datasets using WEKA. International Journal of Computer Applications 54, 21–25. doi:10.5120/8626-2492.
12. Rani, A. S., & Jyothi, S. (2016, March). Performance analysis of classification algorithms under different datasets. In Computing for Sustainable Global Development (INDIACom), 2016 3rd International Conference on (pp. 1584- 1589). IEEE.