Feasible Prediction of Diabetes in Pregnant Woman and Neonatal Mellitus in New Born Child using Machine Learning

M. Shamila[1*], *M.* Srinivas[1], *H.* Aryan Suresh[1], *T.* Viswa Sai Poojith[1], *Dharmendra* Kumar[2], *Bommala* Harikrishna[3]

[1]Department of CSE (AI & ML), GRIET, Hyderabad, Telangana State, India
[2]Uttaranchal Institute of Technology, Uttaranchal University, Dehradun, 248007, India
[3]KG Reddy College of Engineering & Technology, Hyderabad, India

Abstract. Diabetes during pregnancy is a major source of health problems in unborn infants and their moms. Because gestational diabetes can develop to permanent diabetes, ML is an important method for predicting the likelihood of such progression based on the given features. Although the current study may predict lifelong diabetes in pregnant women, it cannot predict the likelihood of neonatal diabetes. As a result, new characteristics are required to improve the forecasting of neonatal mellitus and provide the most accurate and feasible diabetes persistence results in pregnant women. Python scripting and the application of Machine Learning methods such as SVM, KNN, and LR can assist in achieving this aim. The preprocessing ML dataset focusing on Diabetes from the Pima Indian diabetes database collected through Kaggle. In addition, two new attributes were added to the paper's dataset. According to research, machine learning models using characteristics like SVM and decision trees may successfully predict the risk of diabetes in pregnant women. Various factors have been used to predict the beginning of this condition during pregnancy.

1 Introduction

1.1 Introduction

Machine learning algorithms utilize previous data to discover patterns and make predictions about new data. In the case of diabetes prediction, machine learning algorithms may use data from electronic health records, clinical notes, test findings, and patient demographics to identify people at risk of acquiring diabetes. Several researches have looked at the use of ML algorithms to predict diabetes in pregnant women and neonatal diabetes in babies. One research, for example, employed ML algorithms to foresee the risk of gestational diabetes

* Corresponding author: shamila.m@gmail.com

in pregnant women based on demographic and clinical variables. In terms of predicting gestational diabetes, the algorithm has an accuracy of 89.4%.

These algorithms make use of a many of data sources, such as clinical factors, Laboratory tests, and Demographic data. Using these data sets, the algorithms can accurately identify women who are at high risk of having GDM.NDM (neonatal diabetes mellitus) is an uncommon type of diabetes that develops within the first six months of life. Genetic mutations that impair insulin synthesis or function create the disease. NDM is a diverse disorder with different genetic alterations in different people. NDM can be temporary, resolving within a few months of birth, or persistent, necessitating lifetime care.

Gathering relevant data from pregnant women and new-borns, including medical history, glucose levels, BMI, and other factors. This data is preprocessed, selecting important features and ensuring data quality. ML algorithms such as LR, DT, or NN are then trained using this data to establish patterns and relationships between the features and diabetes outcomes. The trained model is evaluated using validation data, and once deemed satisfactory, it can be utilized to predict diabetes risk in pregnant women and neonatal diabetes in new-borns which is very feasible. These predictions aid in identifying high-risk cases and enabling timely interventions and monitoring by healthcare professionals.

1.2 Advantages

1. Early detection allows timely interventions and improves outcomes.
2. Personalized care plans based on individual risk factors enhance treatment effectiveness.
3. Risk stratification optimizes healthcare delivery and resource allocation.
4. Cost savings through reduced complications and hospitalizations.
5. Insights gained contribute to further research and understanding of diabetes.

1.3 Challenges and pitfalls

While the application of ML algorithms for foreseeing diabetes in pregnant women and neonatal diabetes mellitus shows considerable potential, there are various issues that must be solved in order for these models to be accurate and useful. One of the most difficult aspects of forecasting diabetes in pregnant women is the condition's heterogeneity. Gestational diabetes mellitus (GDM) is a complicated illness impacted by a variety of variables such as maternal age, BMI, ethnicity, and diabetes family history. Furthermore, GDM is a changing disease that can occur throughout pregnancy. Because of this complexity, developing reliable prediction models that account for all of these elements and changes is difficult.

2 Existing methods

One frequent strategy is to train machine learning algorithms using clinical data such as glucose levels, blood pressure, and body mass index, as well as demographic and medical history information. These algorithms, which include decision trees, support vector machines, and neural networks, may find patterns and correlations in data to produce diabetes risk predictions. Furthermore, researchers have investigated the use of genetic data and biomarkers to improve prediction accuracy. These techniques are intended to promote early identification and intervention, resulting in improved health outcomes for both pregnant women and new born infants. The paper [8] discussed the role of Intelligent Decision Support Systems (IDSS) in Healthcare Monitoring, especially for heart disease.

Results claimed that IDSS enhances decision-making functionalities in uncertain healthcare scenarios, there by significantly improving the monitoring and remedial activities. The paper [9] explores the distinct ML applications in predicting heart attacks using patient health records. It compares Random Forest and CNN methods, and findings showed that Random Forest's better performance in terms of accuracy. Authors [10] highlighted the significance of ML in prediction, pattern recognition and error reduction across diverse fields, emphasizing the impact of AI in broad domain. Authors [11] suggested data mining techniques to predict disease-prevalence based on symptoms in healthcare data. The appropriate prediction helps healthcare organizations avoid drug shortages and further ensures timely treatment of patients. Authors [12] suggested the different issues and challenges related to IoT based e-healthcare.

Table 1. Summary of approaches.

Ref. No.	Algorithm Used	Results/Remarks
[1]	Random Forest, SVM, Logistic Regression.	Random Forest provided the highest accuracy with 92% compared to other models.
[2]	Random forest, Decision tree, Neural network	The results showed that prediction with random forest could reach the highest accuracy (ACC = 0.8084) when all the attributes were used.
[3]	SVM	Support vector machine classifier demonstrated the highest accuracy, 78.1% predicting the development of gestational diabetes based.
[4]	SVM, Decision tree, Naive Bayes.	The most accurate method was determined to be logistic regression and this resulted in an accuracy of 78.01 percent.
[5]	Random Forest	The overall predictive accuracy of the F1 model was 93.10%, the predictive accuracy of GDM-positive cases 37.10%. The corresponding values for the F2 model 88.70%, and 79.44%.
[6]	The ADA prognostic model, consisting of age and BMI at booking	Selective screening of women for GDM using the ADA model with a risk threshold of 3% gave 93% sensitivity for identification of women with GDM with a 27% reduction in the number of OGTTs required.
[7]	Stepwise Logistic Regression	(age, previous GDM, family history of type 2 diabetes, waist: height and neck: thigh ratios) provided an area under the curve of 0.71 (95%CI 0.68–0.74)

3 Problem statement and objectives

3.1 Problem statement

Diabetes is a chronic condition that affects millions of individuals all over the world and can have major health repercussions, particularly during pregnancy. Gestational Diabetes Mellitus (GDM) is a frequent condition that can result in neonatal mellitus in neonates. Early identification and treatment of GDM are critical for avoiding negative effects for both the mother and the child. Current GDM detection approaches rely heavily on blood glucose tests and clinical risk factors. However, these strategies may not be precise enough to forecast GDM in all circumstances. Machine learning (ML) has showed potential in predicting GDM by analysing massive datasets and finding risk variables that traditional approaches may miss. The problem statement for utilising machine learning to predict diabetes in pregnant women and neonatal mellitus in newborns entails establishing accurate and reliable ML models that gives a feasible prediction of the GDM in pregnant women and neonatal mellitus.

3.2 Objective

The paper's goal is to use machine learning techniques to forecast the risk of permanent diabetes in pregnant women and Neonatal Mellitus mellitus in newborn children. By developing machine learning models and providing a user interface for on-spot data calculation, the paper seeks to facilitate early detection and prevention of diabetes in these populations. Large datasets will be analysed using machine learning algorithms to uncover patterns and risk variables related with persistent diabetes in pregnant women and Neonatal Mellitus diabetes in newborns. By training these models on comprehensive and diverse datasets, the goal is to ensure their reliability and generalizability in real-world scenarios.

4 Proposed method

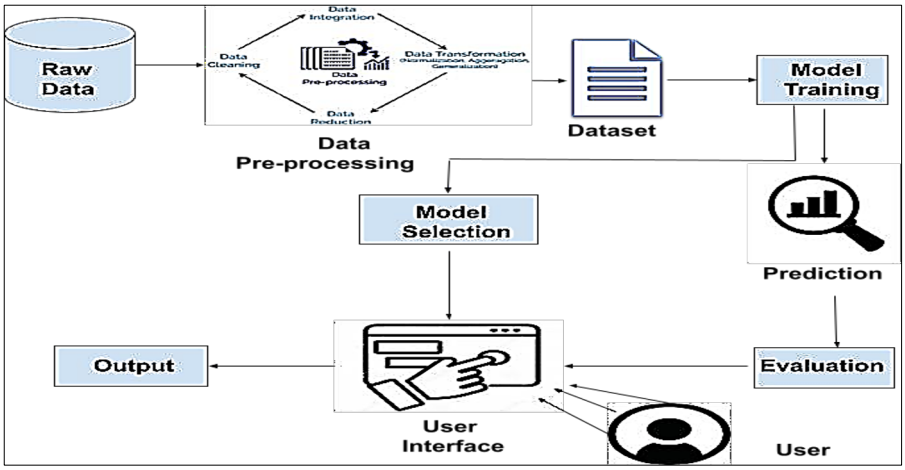

Fig. 1. Diabetes prediction architecture diagram.

The paper aims to develop a feasible predictive model using machine learning techniques to identify the risk of diabetes in pregnant women and neonatal mellitus in new born children. The paper involves collecting comprehensive data, including medical history, glucose levels, BMI, and other relevant factors from pregnant women and new borns. This data is then preprocessed and used to train machine learning algorithms such as logistic regression, decision trees, or neural networks. The trained model will enable early detection of diabetes risk, allowing for timely interventions and personalized care plans. The paper also focuses on risk stratification, cost-effectiveness, and generating valuable insights for further research and understanding of diabetes in these populations.

4.1 Architecture of the proposed work

1. The Raw Dataset serves as the starting point for the architecture. This dataset is specifically curated to contain a collection of data of various females with symptoms of diabetes. The data is typically sourced from various real-world scenarios or collected under specific conditions.
2. Data Preprocessing is an essential step in preparing the blurred images from the RealBlur dataset for training and evaluation. In this step, several operations are performed to transform the raw image data into a suitable format for feeding into a

neural network. The main operations typically include conversion to tensors and resizing

3. After performing various data preprocessing techniques on the raw data set we get a clean dataset that is further used for the paper.
4. Model Training is a Crucial step in the paper as the various models like Naive Bayes, K-Nearest Neighbours, Random Forest, Logistic Regression, SVM are used to train the dataset and the best of the mentioned is used for final execution.
5. Predictions are drawn from the training process to know that which model is better in the prediction.
6. In this step, the model is evaluated using the preprocessed testing data, which consist of diabetes data of females and hence the best model is drawn out of the five.
7. Here the user can input the data he has and then the algorithm runs in the background to calculate whether the user has diabetes or not.
8. Here the algorithm generates output and displays prompts either "You are diabetic" or "You are not diabetic". Along with the accuracy of the calculation.

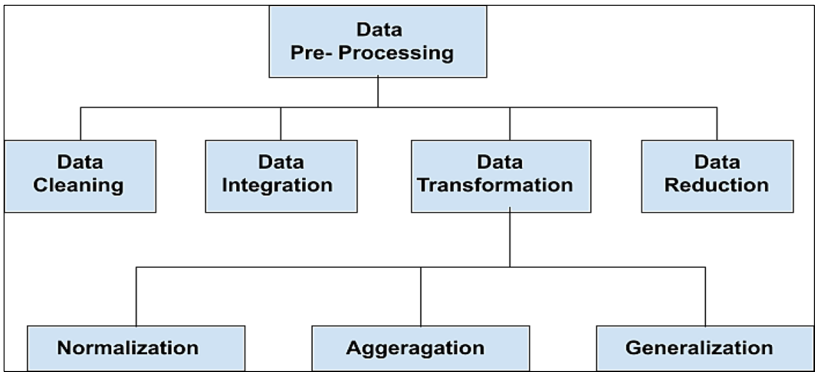

Fig. 2. Data preprocessing diagram.

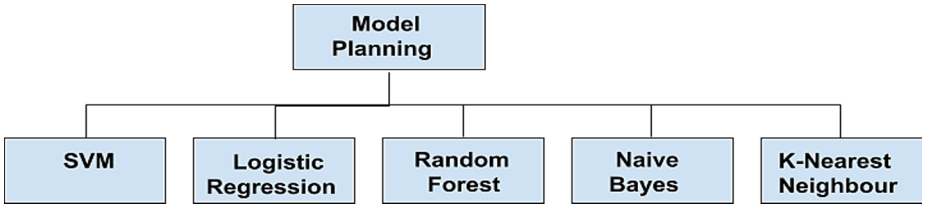

Fig. 3. Model planning diagram.

5 Results and discussions

5.1 Description about dataset

The diabetes dataset consists of various attributes that determine a person's risk of being diagnosed by diabetes. It contains data of various patients bearing one or more of the attributes mentioned in the data set. The detailed meanings of each attribute mentioned in the taken dataset is as follows.

Glucose Level- The concentration of glucose (sugar) in the patient's blood is represented by this aspect. Milligrams per deciliter (mg/dL) or millimoles per litre

(mmol/L) are the most common units of measurement. Elevated glucose levels might be a sign of poor glucose tolerance or diabetes.

Blood Pressure- Typically, this characteristic comprises two values: systolic and diastolic blood pressure. When the heart contracts, systolic blood pressure indicates the pressure in the arteries, whereas diastolic blood pressure shows the pressure when the heart is at rest. Hypertension (high blood pressure) is a risk factor for gestational diabetes.

BMI (Body Mass Index)- BMI is a measure of body fat that is dependent on a person's height and weight. It is computed by dividing the weight in kilogrammes by the height in metres squared. BMI indicates if a patient is underweight, normal weight, overweight, or obese. Obesity is a major risk factor for developing gestational diabetes.

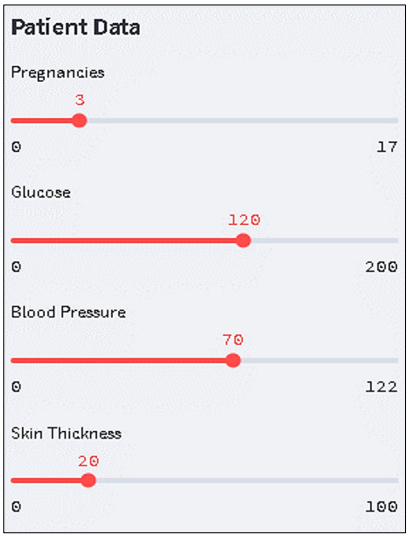

Fig. 4. Input slider-1.

5.2 Experimental results

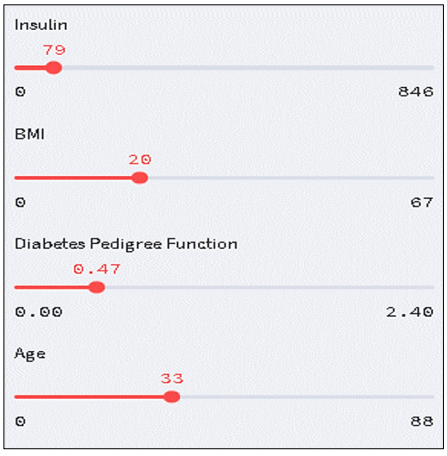

Fig. 5. Input slider-2.

Out[3]:

	Number of times pregnant	Plasma glucose concentration	Diastolic blood pressure	Triceps skin fold thickness	2-Hour serum insulin	Body mass index	Diabetes pedigree function	Age (years)	Class variable
0	6	148	72	35	0	33.6	0.627	50	YES
1	1	85	66	29	0	26.6	0.351	31	NO
2	8	183	64	0	0	23.3	0.672	32	YES
3	1	89	66	23	94	28.1	0.167	21	NO
4	0	137	40	35	168	43.1	2.288	33	YES
...	...	...	...	...	...	...	...	...	...
763	10	101	76	48	180	32.9	0.171	63	NO
764	2	122	70	27	0	36.8	0.340	27	NO
765	5	121	72	23	112	26.2	0.245	30	NO
766	1	126	60	0	0	30.1	0.349	47	YES
767	1	93	70	31	0	30.4	0.315	23	NO

768 rows x 9 columns

Fig. 6. Training data stats-2.

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age
0	3	120	70	20	79	20	0.47	33

Fig. 7. Patient data.

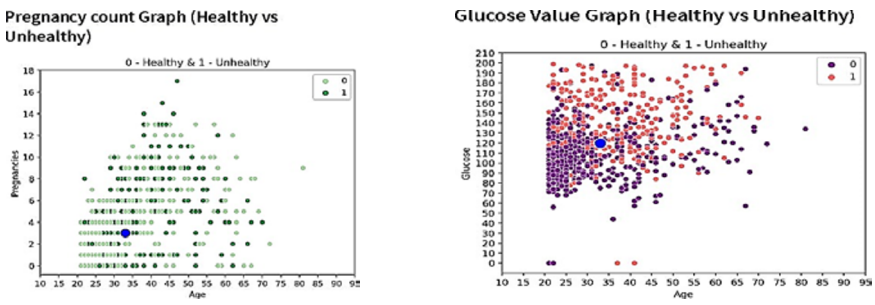

Fig. 8. Pregnancy count graph & glucose value graph.

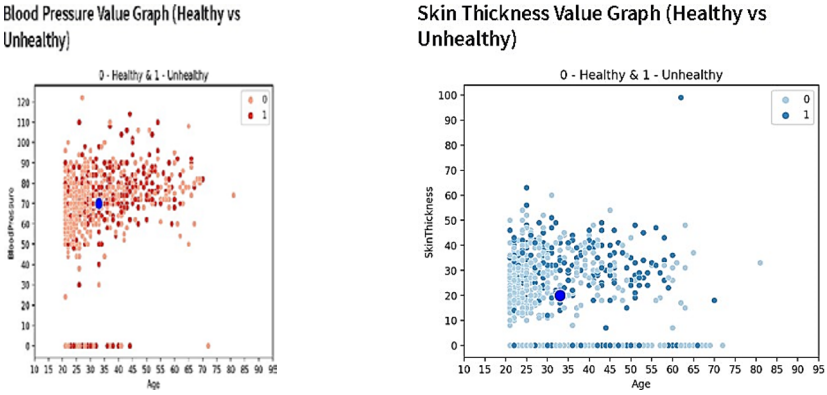

Fig. 9. Blood pressure value graph & skin thickness value graph.

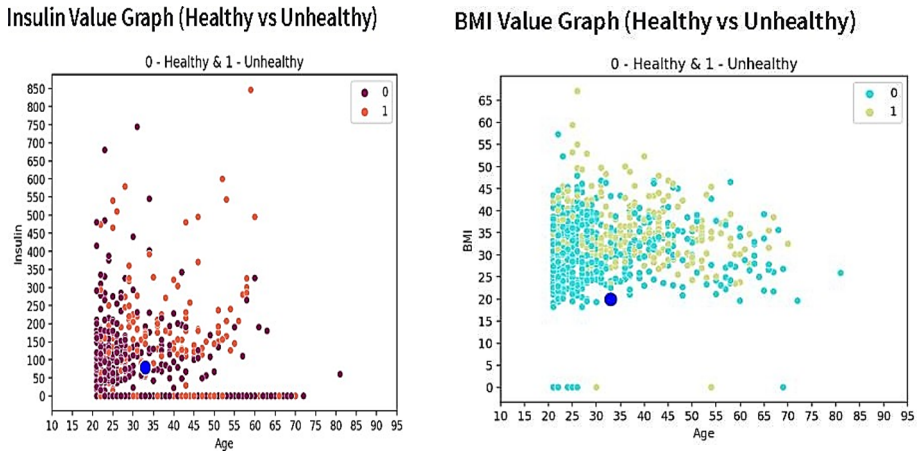

Fig. 10. Insulin value graph & BMI value graph.

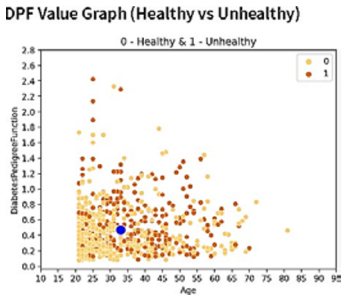

Fig. 11. DPF value graph.

Fig. 12. Output and accuracy.

The proposed method of predicting diabetes in pregnant women using machine learning holds significant potential in advancing healthcare outcomes in this domain which is feasible. By leveraging machine learning algorithms and predictive modelling techniques, this approach aims to provide a feasible prediction of diabets and risk assessment, enabling timely interventions and personalized care. One of the key advantages of this method is its potential to improve maternal and neonatal health outcomes by identifying high-risk cases accurately. Machine learning models can analyze large amounts of data from pregnant women, including medical history, lifestyle factors, and various clinical parameters, to identify patterns and generate predictive insights.

6 Conclusion and future enhancements

Feasible prediction of diabetes in pregnant women and neonatal mellitus in new born children using machine learning revolves around developing predictive models to predict the risk of gestational diabetes in women expectant mothers and the likelihood of neonatal diabetes in new born. By leveraging machine learning algorithms and analysing various factors such as medical history, clinical parameters, lifestyle, and genetic predisposition, these models aim to provide early detection and personalized risk assessment. The goal is to enable healthcare professionals to intervene proactively, implement preventive measures. The potential of machine learning to predict gestational diabetes and neonatal mellitus is promising and encompasses several potential advancements and areas of exploration. Here

are some key aspects that offer significant opportunities for further development in terms of accuracy and longitudinal analysis.

References

1. S. Saru, S. Subashree, Intl. J. Emerg. Tech. Innova. Engg **5**, 4 (2019)
2. Q. Zou, K. Qu, Y. Luo, D. Yin, Y. Ju, J. Zhang, Sec. Computa. Genom **9**, (2018)
3. G. Cubillos, M. Monckeberg, A. Plaza, M. Morgan, Pregn. Childbirth, **469** (2021)
4. Z. Zaman, A. Shohas, M. H. Bijoy, M. Hossain, S. A. Sakib, Intl. J. Advanc. Life Sci. Res **5**, 1 (2022)
5. Lili Wei , Yueshuai Pan , Yan Zhang, Futur. Oncolo **8**, (2021)
6. S. Amarnath, M. Selvamani, V. Varadarajan, Sci. Adv. Mater **33**, 9 (2021)
7. S. L. White, D. A. Lawlor, A. Briley, PLoS ONE, **11**,12 (2016)
8. R. P. Ram Kumar, R. Tabassum, Intl. J. Creat. Res. Thoug **6**, 1 (2018)
9. R. P. Ram Kumar, S. Polepaka, Performance Comparison of Random Forest Classifier and Convolution Neural Network in Predicting Heart Diseases, in Proceedings of the Third International Conference on Computational Intelligence and Informatics, (eds) K. Raju, A. Govardhan, B. Rani, R. Sridevi, M. Murty, Advances in Intelligent Systems and Computing, vol 1090. Springer, Singapore (2020)
10. R. P. Ram Kumar, P. Sanjeeva, S. F. Lazarus, D. V. Krishna, Intl. J. Inno. Tech. Explor. Engg **8**, 11S2 (2019)
11. A. Sankaridevi R. P. Ram Kumar, R. Jayakumar, Intl J. Recen. Tech. Engg **7**, 5C (2019)
12. M. Shamila, K. Vinuthna, A. K. Tyagi, *A Review on Several Critical Issues and Challenges in IoT based e-Healthcare System*, in Proceedings of the 2019 International Conference on Intelligent Computing and Control Systems (ICCS), Madurai, India, (2019)