Early Gestational Diabetes Mellitus Diagnosis Using Classification Algorithms: an Ensemble Approach

Oluwafemi O. S. Abe, Olumide O. Obe, Olutayo K. Boyinbode and Biodun N. Olagbuji

September 2, 2023

Early Gestational Diabetes Mellitus Diagnosis Using Classification Algorithms: An Ensemble Approach

Oluwafemi O. S. Abe
Department of Computer Science,
The Federal University of Technology,
Akure, Nigeria.
osabe@futa.edu.ng

Olumide O.Obe
Department of Computer Science,
The Federal University of Technology,
Akure, Nigeria.
ooobe@futa.edu.ng

Olutayo K. Boyinbode
Department of Information Technology,
The Federal University of Technology,
Akure, Nigeria.
okboyinbode@futa.edu.ng

N. Biodun Olagbuji
Department of Obstetrics and Gynecology,
Ekiti State University,
Ado-Ekiti,
Nigeria.
ebiodun_olagbuji@yahoo.com

***Abstract* - In the next twenty years, Type 2 diabetes may affect over 50% of GDM patients, and infants and adults can acquire the disease. It is critical to consider both the mother's and the children's short-term and long-term challenges. In the current situation, early diagnosis is essential due to maternal morbidity and mortality and fetal problems. Early identification and prevention are inefficient and often problematic in developing and underdeveloped nations. Deciding GDM needs a well-designed approach, which is urgently needed. This study's primary aim is to forecast GDM in the first trimester. To figure out if a pregnant woman is at risk of GDM or not. This research used KNN, LR, and RF for classification and an ensemble (majority vote) model. After an error-free validation, the data was passed to the machine learning pre-trained model file, which in turn returns the predicted value to the frontend-designed with HTML and CSS. Python serving framework was used to connect the frontend code with the model. The ML model file, web codes, and flask codes were uploaded to the Github repository in preparation for final deployment on the server. The codes on Github were connected to Heroku where the web application is hosted. The user's interface with the web application to access the pre-trained model to make predictions.**

Keywords— algorithms, classifier, diabetes, diagnosis, ensemble.

I. INTRODUCTION

In June 2022, 6.64 billion people had mobile devices, while 650 million mobile users are in Africa [1]. By 2025, the number of people using mobile phones in the world was projected to be 7.49 billion [1], [2]. These devices' availability will go a long way to help diagnose some deadliest diseases like Diabetes Mellitus (DM). Gestational Diabetes Mellitus (GDM) as a type of DM refers to glucose intolerance which is first observed in pregnant women and this may affect the woman's health postpartum [1]. Early prediction and classification of diabetes are large [1] to reduce the threat to the life of expectant mothers and their babies which was estimated to cause the death of 629 million by the end of 2045 [2], [3].

GDM is commonly diagnosed between 24-28 weeks of pregnancy. However, earlier discovery is preferable since it may avoid or minimize the risk of unfavourable pregnancy outcomes[4] or the risk of developing Type 1 Diabetes Mellitus in future [5]. Currently, it has become a significant public health concern due to its long-term complication. Since most pregnant women do not have full knowledge of the complications associated with GDM, they never embark on early diagnosis and treatment [6]. The application of Machine learning will reduce the cost and speed up the process [7] of predicting the GDM disease.

Research on medical diagnosis has undergone many various phases ranging from Statistical Methods, Bayesian Inference, Utility Theory, Boolean Logic, discriminant analysis etc [8]. When it is evident that more complex cases cannot be adequately handled with statistical tools, Artificial Intelligence (AI) principles were applied[3], [8]. Even though medicine is generating a huge amount of data every day locally, little has been done to collate and use the data to solve the challenges that face a successful interpretation of medical examination results. Several data mining approaches [6], [9]have been applied to GDM [3] but limited research was conducted using the Majority Vote Ensemble approach with mobile diagnosis application. This paper starts by applying data pre-processing on the GDM dataset collected from the Strategy for Comprehensive Gestational Diabetes Control (SCGDC) Project Database, Cleaning of the data, training, validation, testing and implementation. The classifiers were evaluated using metrics like accuracy, sensitivity and specificity.

Related Work on Predictive Models of GDM

"Machine learning risk score for the prediction of gestational diabetes in early pregnancy in Tianjin, China" was presented by [10]. The XGBoost model was implemented using a free, publicly available software interface, and 1484 (7.6%) of the women have GDM. Risk variables were Body Mass Index (BMI), maternal age, Fasting Plasma Glucose (FPG) at registration, and Alanine AminoTransferase (ALT). The XGBoost model outperformed the logistic model in terms of AUR (0.742 vs. 0.663, P <0.001). Additional GDM risk variables should be considered for improved prediction. Reference [11] authored a paper on "the comparison of

Machine Learning Methods and Conventional Logistic Regressions for Predicting Gestational Diabetes Using Routine Clinical Data: A Retrospective Cohort Study". There were 22,242 singleton pregnancies included in the study, and 3182 (14.31%) had GDM. GDBT, AdaBoost, LGB, Logistic, Vote, XGB, DT, and RF variables were trained) and (stepwise logistic regression and logistic regression with RCS). The machine learning and logistic regression models performed well on the validation dataset (Area Under Curve 0.59 - 0.74). Overall, the GBDT model outperformed the other machine learning approaches (AUC 0.74, 95% CI 0.71-0.76), with only minor differences.

In [12] "Prediction of Gestational Diabetes based on nationwide electronic health records was conducted." The study used countrywide electronic health data to predict gestational diabetes. The models predict GDM with high accuracy even at pregnancy initiation (area under the receiver operating curve (auROC) = 0.85), outperforming a baseline risk score (auROC = 0.68). Added populations are needed to assess the real-world clinical utility of the model. [13] worked on "Ensemble Classifier Technique to Predict Gestational Diabetes Mellitus (GDM)." The ensemble model was then used, together with a voting classifier, to find the proper class labels for the applied data examples. Extensive experiments on many aspects were carried out to confirm the competency of the offered models. From the results of experimental analysis, the ensemble model outperformed the classical ML models, and it achieved a precision of 94%, recall of 94% and F-score of 94%. Reference [10] published a paper titled "Prediction Method of Gestational Diabetes Based on Electronic Medical Record Data." The training set was used to build a logistic regression model, and the test set data were fed into the prediction model for prediction. When the Weight Base Feature Selection (WBFS)-filtered features are included, the accuracy, F1 value, and AUC value of logistic regression are 0.809, 0.881, and 0.825, respectively, an improvement of 12% over when the feature is not employed. The findings show that using an electronic medical record data drive can significantly increase the accuracy of forecasting gestational diabetes.

II. METHODOLOGY

The research used the ensemble method (Majority Voting) for gestational mellitus prediction. The methods acquired information from the Gestational Diabetes Mellitus dataset. To enhance the performance of the ensemble methods. The method combines predictions from a set of different supervised classification algorithms K-Nearest Neighbor (KNN), Random Forest (RF), and Logistic Regression (LR) Algorithms to improve prediction accuracy. Every classifier in the majority voting ensemble votes for a certain class label, and the final output class label obtains more than half of the votes; otherwise, a rejection choice is provided. The ensemble model was implemented in a Python environment.

A. System Design and Implementation

The architecture of the ensemble predictive models for the gestational diabetes mellitus disease is shown in Fig. 1. The collected data was cleaned, selected, integrated, and discretized. The features were trained using classification algorithms called based models: KNN, Random Forest, and LR classifiers for the model's construction. These were used as the base learners/classifiers. The Majority Voting ensemble was another component used to combine the predictions of the three-based models and train. Classification evaluation metrics such as accuracy, confusion matrix, recall and precision, FI score and mean absolute error were considered to measure the final prediction. The predictive results show the outcome of the GDM diagnosis model.

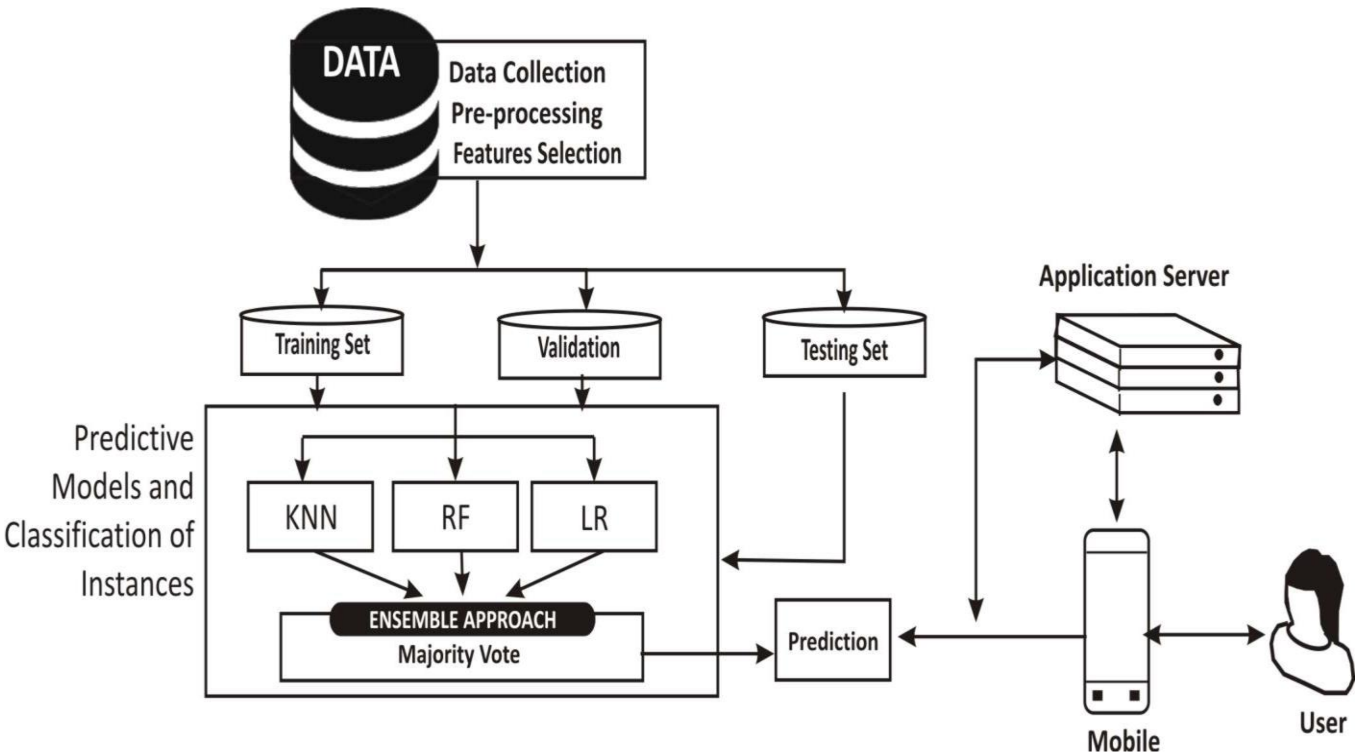

Fig. 1. The Architecture of the Diagnosis GDM

B. Data Collection

The Gestational Diabetes dataset used in this research work was obtained from experimental data sources from State Government Hospitals, in Ondo State, Nigeria. The datasets have 12,582 clinical instances. Each clinical instance holds 12 attributes and one target attribute. The target attributes refer to the status of the presence of GDM in the patient. It is represented by an integer value "0" or "1" where "0" signifies absence and the value "1" signifies the presence of GDM. Table I shows the parameters used from the dataset and observed symptoms.

TABLE I: ABSTRACT FROM DATASET AND SYMPTOMS DETAILS

Attribute	**Range**	**Data Type**
Age	15 to 50	Numeric
Weight (kg)	30 - 130	Numeric
Height (meters)	1 - 2.5	Numeric
Pregnancy State	Yes/No	Nominal
Week of Pregnancy	23 to 24	Numeric
Parity (6 months)	1 to 10	Numeric
Number of previous pregnancies	1 to 20	Numeric
FPG (m/l)	4.2 - 5.5	Float
Gestational Diabetes	0 or 1	Nominal
Feeling usual thirst	Yes/No	Nominal
Frequent Urination	Yes/No	Nominal
Experience Blurred Vision	Yes/No	Nominal
Experience Fatigue	Yes/No	Nominal

- *Data Pre-Processing:* The data selection was done by retrieving data from a database relevant to the GDM. This was followed by Data Integration preprocessing technique [14] that merges the data from multiple data sources (hospitals) into a coherent data store. The features of GDM are parity, BMI, number of pregnancies, FPG, etc. represented in the form: $y_1, y_2, \ldots y_n \rightarrow z$, Data sources (DS) contained the form in (1) and (2).

$$DS_n = \{ y_1, y_2, \ldots y_n \rightarrow z_j \} \quad (1)$$

By merging different data sources as one coherent dataset, we form (2).

$$D_{int} = DS_a \cup DS_b \ldots \cup DS_n \quad (2)$$

Where D_{int} is the data integration; ∪ is the union to merge data sources, DS_a, DS_b DS_n are the data sources for various centres and *j* represents the number of classes.

- *Data Cleaning* - The data was cleaned at this step by using the Imputation (zero values) and Case Deletion techniques.
- *Data Discretization* as suggested by [15] was used to minimize the number of continuous feature values since a high number of potential feature values adds to the sluggish and ineffective machine learning process. In this work, data discretization was done by converting nominal input values (attributes) into numeric values [16]. After all the processes, the attribute was reduced to 6 and 5,981 clinical instances. Table 2 shows the extract from the GDM Dataset. The Min-Max Data normalization technique was applied to get the input values normalized. The Min-Max normalization equation is given in (3).

$$x_{inew} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (3)$$

Where
x_{inew} = new value for the variable with value *x*
x = current value for the variable with value *x* (old value of individual feature being scaling)
x_{min} = minimum value in the dataset
x_{max} = maximum value (GDM dataset). To illustrate the normalization of the BMI using Table II.

Table II: EXTRACT FROM THE GDM DATASET

Age	**BMI**	**NM**	**NP**	**FPG**	**Outcome**
27	68.75	1	2	3.43	0
28	68.75	0	0	4.2	0
35	67.52	1	4	4.29	0
25	65.36	0	0	5.2	1
34	64.06	0	2	6	1
25	63.06	2	2	4.7	0
32	63.06	0	2	6.3	1
26	60.97	0	0	3.4	0

NM - Number of Miscarriage, NP – Number of Pregnancy

Equation (5) gives, x_{inew} = ?, x = 60.97, x_{min} = 13.14878893, x_{max} = 68.75

$$x_{inew} = \frac{60.97 - 13.149}{68.75 - 13.149}$$

$$x_{inew} = \frac{47.821}{55.601}$$

$$x_{inew} = 0.86$$

The Normalization of BMI 60.97 is 0.86, the same process was used for other features.

C. Classifiers

- *Random Forest (RF)*

The Random Forest was constructed from the training GDM dataset by splitting it into feature subsets using a greedy algorithm. The features such as age, parity, number of pregnancies, number of miscarriages, FPG etc. Fig. 2 was used to illustrate the process.

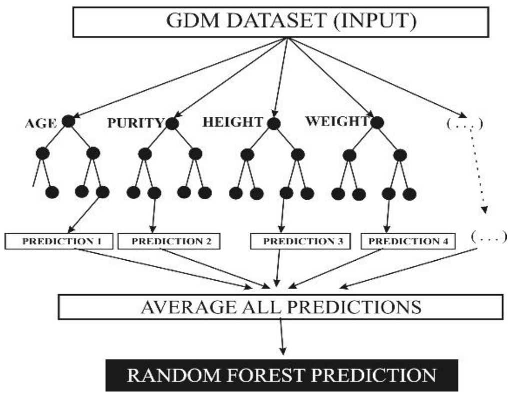

Fig. 2. Application of the Random Forest

- *K-Nearest Neighbor (KNN)*

"The k-Nearest Neighbor algorithm is a machine learning algorithm" as stated by [18], [19]. The goal is to learn the labels of the instances in the GDM training set, and then predict the label of each new case using those labels. The rationale behind such a method assumes that the features that are used to describe the domain [20] "points are relevant to their labelling in a way that makes close-by points likely to have the same label." A k-nearest-neighbour classifier as expressed in (4) looks for the k-training tuples that are most like an unknown tuple when it is given an unknown tuple [21]. These k-training tuples are the k "nearest neighbours" of the unknown tuple. "Closeness" is defined in terms of a distance metric, such as Euclidean distance [21]. The Euclidean distance between two tuples [22], $X_1 = (x_{11}, x_{12}, \ldots, x_{1n})$ *and* $X_2 = (x_{21}, x_{22}, \ldots, x_{2n})$, is

$$dist\ (X_1, X_2) = \sqrt{\sum_{i=1}^{n}(x_{1i} - x_{2i})^2} \quad (4)$$

- *Logistic Regression (LR)*

As defined by [23] "Logistic regression as a type of classifier was used for the prediction of the outcome of a categorical dependent variable from a set of predictor or independent variables" [24]. It calculates the probability of a discrete outcome based on the GDM data. It was based on linear regression. The logit function is applied to the GDM output of linear regression to find the probability. A logit function is formulated in (5) and the logit function can convert any real value to a value between 0 and 1.

$$logit\ (\hat{z}) = \frac{1}{(1+e^{-2})} = \frac{e^2}{(e^2+1)} \quad (5)$$

The variable $\hat{z}$ is a measure of the total contribution of all the independent variables used in the model and is known as the logit.

- *Ensemble - Majority (Hard) Voting*

Voting is a mechanism for combining the decisions of many classifiers [25]. The concept divided the GDM training data into smaller equal sections and developed a classifier for each subset. The most basic kind of voting is plurality or majority voting, in which each classifier gives a single vote. Most votes were used to make the final judgment; so, the class with the most votes was the final prediction. The ultimate selection is made by tallying all votes and selecting the class with the highest aggregate. Fig. 3 depicts the architecture of the voting ensemble model.

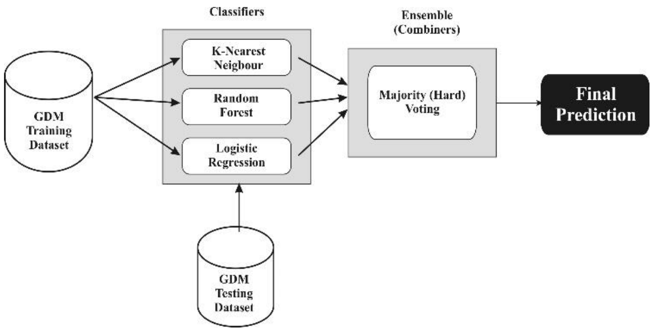

Fig. 3. Architecture of GDM Majority Voting (Ensemble)

In the ensemble technique, the final class label will be predicted as the class label that has been predicted most frequently by different classification models [8].

Given a GDM dataset D with instances N and C as the class label as shown in (6).

$$D = \{(y_n, x_n), n = 1, \ldots, N\} \quad (6)$$

where y_n is the target class; and x_n stands for feature vectors of the nth instance. Also, define a set of classifiers $S = \{M_1, M_2, M_3\}$. Each instance $x \in D$ is assigned to have one of the C classes. Each classifier has its prediction for each instance. The final class allotted to each instance is the class forecast by most classifiers (gaining the majority votes) for that instance. This is formulated as follows.

Let $c_l \in C$ denote the class of an instance x predicted by a classifier M_1, and let a counting function F_k be defined in (7) as:

$$F_k(c_i) = \begin{cases} 1 & c_i = c_k \\ 0 & c_i \neq c_k \end{cases} \quad (7)$$

where c_l and c_k are the classes of C. The count of total votes T_k for class c_k can then be defined in (8) as:

$$T_k = \sum_{l=1}^{M} F_k(c_i) \quad (8)$$

The predicted class C for an example x using the classifier set S is defined to be a class that gains the majority vote as defined by (9) as:

$$C = (x)\, T_k \quad (9)$$

The Frontend development is a web page designed compressed to the mobile device with Hypertext Makeup Language (HTML) and Cascading Style Sheet (CSS) as shown in Fig. 4.

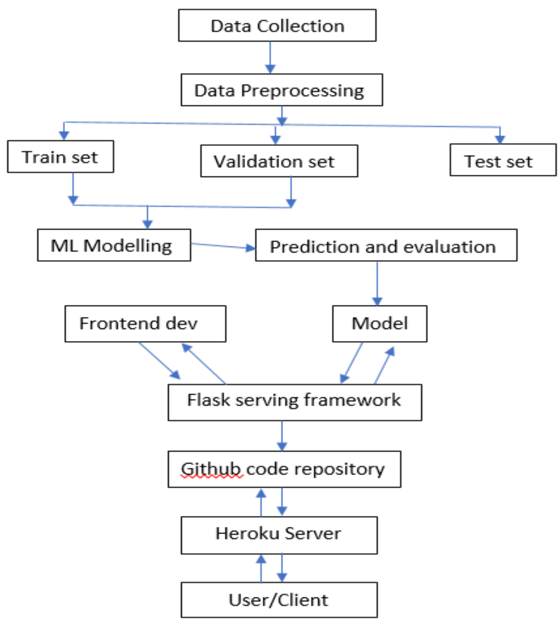

Fig. 4. Link within the Stages of Model

The model was saved into a pickle file for deployment, that way, the training, and modelling were not done online. The pickle file was the pre-trained model that will do the prediction. Python serving framework – Flask, was used to connect the frontend code with the model. The user input validation was done in this phase. The ML model file, web codes, and flask codes were uploaded to the Github repository to prepare for final deployment on the server. The codes on the Github repo were connected to Heroku where the web app is hosted.

III. EVALUATION

Performance measures were used to compare the results from all the classifiers and the ensemble. The measures include sensitivity, specificity, F-1 Score, and accuracy. The ensemble methodologies used in the article prove the efficiency of multi-classifiers in combining base-leaner forecasts with superior classification results than individual-based classifier performances. The ensemble has 98.45% sensitivity, 85.06 % Specificity and 91.72% accuracy. Table III and Fig. 5 illustrate the details of the performance.

TABLE 3: THE PERFORMANCE MEASURE OF THE MODEL

Model / Evaluation (%)	**Sensitivity**	**Specificity**	**F-1 Score**	**Accuracy**
LR	96.958	81.804	90.040	89.34
KNN	89.259	76.668	83.864	82.93
R.Forest	97.655	41.904	76.165	69.62
MV Ensemble	98.447	85.061	92.196	91.72

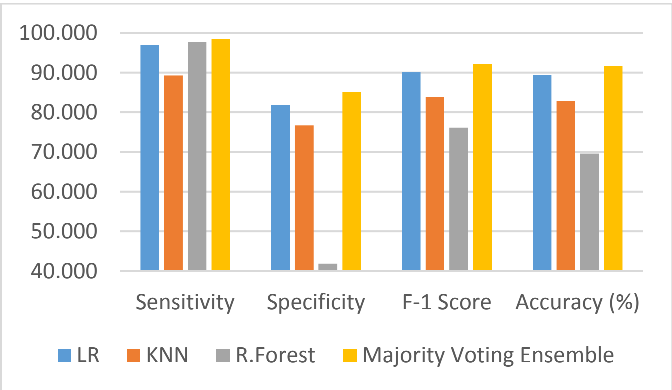

Fig. 5. The Chat of the Performance Measure of the Model

- *Interfaces and System Testing*

  When the user inputs the required data, the data is validated for correctness. After an error-free validation, the data of the user (Age, height, weight, number of pregnancies held before now, and parity). Also, the symptoms will the ticked. The data is passed to the machine learning pre-trained model file which in turn returns the predicted value to the frontend. The output of the prediction is displayed on the client's mobile device. Fig. 5 shows the mobile view of validation user data computation and predicted outcomes.

Fig. 6. The Application View of Validation User Data Computation and Predicted Outcomes

IV. DISCUSSION

Nowadays, gestational diabetes is a common metabolic condition among pregnant women. This disease is linked to several risk factors, which might cause issues for the mother and the newborn. If untreated, there may be long-term risks for both the mother and the child. Therefore, it is important to use the right screening and diagnosis methods to maintain optimal glycemic management. Although glucose tolerance normally recovers to normal shortly after birth, convincing evidence suggests that women with GDM have a high lifetime risk of developing diabetes. As a result, GDM medications are often less successful in treating women who acquire GDM and do not have these prevalent risk factors before the second trimester. This study unequivocally proves that when looking at the inputs to the models, there is at least one input value for which the patient should seek medical assistance from a staff member of a hospital. Our prospective and multicenter study is the first clinical examination that supports the GDM diagnosis for pregnant women in resource-limited settings, using age, parity, height, weight, and the number of pregnancies, as well as symptoms of GDM. A mobile device (such as a smartphone) with internet access can be used to forecast the FPG and decide if the person is at risk. Our research has shown that the ensemble technique may provide a precise diagnosis with lower operating expenses and more effectiveness. Our study shows that our app has a bright future in advancing precision medicine, long-distance healthcare, and maternal health for expectant women.

Through this study, every expectant mother has the chance to decide her risk before ever visiting the hospital. Since GDM is so common, many expectant mothers worry about getting it. If GDM is discovered early, the risk during pregnancy is lower. To reduce complications, it is recommended that women who are at risk for GDM be identified early. If the facilities and human resources are scarce, this research may also be used in hospitals and clinics as a method to check for GDM. AI applications for illness prediction are becoming more popular, and they perform better in terms of making medical decisions. We recommend that future research broaden the dataset's coverage and repeat the procedure to confirm the efficiency of the AI algorithms.

REFERENCES

[1] D. Vigneswari, N. K. Kumar, V. G. Raj, A. Gugan, and S. R. Vikash, "Machine Learning Tree Classifiers in Predicting Diabetes Mellitus," *2019 5th Int. Conf. Adv. Comput. Commun. Syst.*, pp. 84–87, 2019.

[2] B. P. Nguyen *et al.*, "Computer Methods and Programs in Biomedicine Predicting the Onset of Type 2 diabetes using wide and deep learning with electronic health records," *Comput. Methods Programs Biomed.*, vol. 182, p. 105055, 2019, doi: 10.1016/j.cmpb.2019.105055.

[3] O. S. Abe, O. O. Obe, O. K. Boyinbode, and O. N. Biodun, "Classifier Algorithms and Ensemble Models for Diabetes Mellitus Prediction: A Review," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 10, no. 1, pp. 430–439, 2021, doi: 10.30534/ijatcse/2021/641012021.

[4] R. Akula, N. Nguyen, and I. Garibay, "Supervised Machine Learning based Ensemble Model for Accurate Prediction of Type 2 Diabetes," *Conf. Proc. - IEEE SOUTHEASTCON*, vol. 2019-April, 2019, doi: 10.1109/SoutheastCon42311.2019.9020358.

[5] A. Katsarou *et al.*, "Type 1 Diabetes Mellitus," *Nat. Publ. Gr.*, vol. 3, no. March, pp. 1–18, 2017, doi: 10.1038/nrdp.2017.16.

[6] E. C. Igodan, O. O. Obe, A. F. Thompson, and O. Owolafe, "A Hybrid Feature Ensemble Method for Cervical Cancer Classification," *10th IEEE Int. Conf. E-Health Bioeng.*, pp. 0–4, 2022.

[7] M. I. Oladunjoye and O. O. Obe, "Deep Neural Networks in the Discovery of Novel Antibiotics Drug Molecule : A Review," vol. V, no. Ix, pp. 147–150, 2020.

[8] B. O. Afeni, "Development Of Ensemble Predictive Models For Coronary Heart Disease (CHD," Federal University of Technology, Akure., 2019.

[9] F. O. Akinloye, O. Obe, and O. Boyinbode, "Development of an affective-based e-healthcare system for autistic children," *Sci. African*, vol. 9, p. e00514, 2020, doi: 10.1016/j.sciaf.2020.e00514.

[10] H. Liu *et al.*, "Machine learning risk score for prediction of gestational diabetes in early pregnancy in Tianjin, China," *Diabetes. Metab. Res. Rev.*, pp. 0–3, 2020, doi: 10.1002/dmrr.3397.

[11] Y. Ye, Y. Xiong, Q. Zhou, J. Wu, X. Li, and X. Xiao, "Comparison of Machine Learning Methods and Conventional Logistic Regressions for Predicting Gestational Diabetes Using Routine Clinical Data: A Retrospective Cohort Study," *J. Diabetes Res.*, vol. 2020, 2020, [Online]. Available: https://www.hindawi.com/ journals/jdr/2020/4168340/.

[12] N. S. Artzi *et al.*, "FOCUS | Letters FOCUS | Letters Prediction of gestational diabetes based on nationwide electronic health records Letters | FOCUS," *Nat. Med.*, vol. 26, no. January 2020, doi: 10.1038/s41591-019-0724-8.

[13] A. Sumathi and S. Meganathan, "Ensemble classifier technique to predict gestational diabetes mellitus (GDM)," *Comput. Syst. Sci. Eng.*, vol. 40, no. 1, pp. 313–325, 2022, doi: 10.32604/CSSE.2022.017484.

[14] O. C. Olayemi, O. O. O., B. A. Ojokoh, and A. I. Peter, "Comparative Analysis of Predictive Models for Diagnosis of Lower Respiratory Infections among Paediatric patients," *Comput. Rev. J.*, vol. 8, 2020.

[15] Khouloud-Abdel Aziz Safi E., "Predicting Hypoglycemia In Diabetic Patients Using Machine Learning Techniques," Faculty of the American University of Sharjah, 2014.

[16] U. A. Zia and N. Khan, "Predicting Diabetes in Medical Datasets Using Machine Learning Techniques," *Int. J. Sci. Eng. Res.*, vol. 8, no. 5, pp. 1538–1551, 2017.

[17] O. O. O. O.E. Oduntan, I.A. Adeyanju, A.S. Falohun, "A comparative analysis of Euclidean distance and cosine similarity measure for automated essay-type grading," *J. Eng. Appl. Sci.*, vol. 13, no. 11, pp. 4198–4204, 2018.

[18] M. Abed and T. Ibrikci, "Comparison between Machine Learning Algorithms in the Predicting the Onset of Diabetes," *2019 Int. Artif. Intell. Data Process. Symp.*, pp. 1–5.

[19] S. K. Dey, A. Hossain, and M. M. Rahman, "Implementation of a Web Application to Predict Diabetes Disease: An Approach Using Machine Learning Algorithm," *2018 21st Int. Conf. Comput. Inf. Technol. ICCIT 2018*, pp. 1–5, 2019, doi: 10.1109/ICCITECHN.2018.8631968.

[20] M. A. Sarwar, N. Kamal, W. Hamid, and M. A. Shah, "Prediction of Diabetes Using Machine Learning Algorithms in Healthcare," *2018 24th Int. Conf. Autom. Comput.*, no. September, pp. 1–6, 2018, doi: 10.23919/IConAC.2018.8748992.

[21] R. E. Izzaty, B. Astuti, and N. Cholimah, *Data Mining Concepts and Techniques*, 3rd ed. USA: Morgan Kaufmann Publishers, 2012.

[22] D. R. Parikh, Y. R. Bhagat, and N. R. Ghanwat, "Prediction of Probability of Chronic Diseases and Providing Relative Real-Time Statistical Report using data mining and machine learning techniques," vol. 5, no. 4, 2016.

[23] T. Witelski, *Methods of Mathematical Modelling*, 1st ed. Switzerland: Springer International Publishing, 2015.

[24] S. Kramer and C. Helma, *Introduction to Artificial Intelligence*, 2nd ed. United Kingdom: Springer International Publishing, 2017.

[25] J. L. Fernandez-Aleman, J. M. Carrillo-De-Gea, M. Hosni, A. Idri, and G. Garcia-Mateos, "Homogeneous and heterogeneous ensemble classification methods in diabetes disease: A review," *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, pp. 3956–3959, 2019, doi: 10.1109/EMBC.2019.8856341.

[26] F. T. Matthew *et al.*, "Development of mobile-interfaced machine learning-based predictive models for improving students' performance in programming courses," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 5, pp. 105–115, 2018, doi: 10.14569/IJACSA. 2018.090514.