ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 13, Issue 4, pp: (1-11), Month: October - December 2025, Available at: www.researchpublish.com

Machine Learning-Based Extreme Gradient Boost Algorithm for the Prediction of Gestational Diabetes Mellitus among Pregnant Women

Ogechukwu Emmanuella Ndukwe[1], Laeticia Nneka Onyejegbu[2], Friday E. Onuodu[3]

[1,2,3]Department of Computer Science, University of Port Harcourt, Rivers State, Nigeria

DOI: https://doi.org/10.5281/zenodo.17265472

Published Date: 04-October-2025

Abstract: **One of the primary information technologies for antenatal management is the utilization of a machine learning system for healthcare interventions. A machine learning tool is a vital resource for antenatal care and the monitoring of pregnant women, offering services that support healthcare providers, pregnant women, and their family members in predicting and providing adequate healthcare. One of the major challenges faced by pregnant women is the issue of gestational diabetes during pregnancy. It comes with complications such as spontaneous abortion, secondary infections, fetal malformation, hypertensive disorder, cholestasis, obstructed vaginal delivery, and kidney disease. This study aims to develop a machine learning model for predicting gestational diabetes mellitus (GDM) using medical history and socioeconomic features of pregnant women, employing three machine learning models: Extreme Gradient Boosting (XGBoost), Random Forest (RF), and Support Vector Machine (SVM). The results of the implementation show that XGBoost achieved an F-score of 0.6234 and an accuracy of 83.11%, RF achieved an F-score of 0.6053 and an accuracy of 82.79%, while SVM obtained a performance accuracy of 83.44% with an F-score of 0.5731. The developed machine learning approach will guide healthcare practitioners on the deployment of machine learning models for gestational diabetes mellitus prediction.**

Keywords: **machine learning, disease prediction, gestational diabetes mellitus, Extreme Gradient Boosting, Artificial Intelligence.**

I. INTRODUCTION

According to the World Health Organization (WHO, 2020), the African Region lacked 817,992 healthcare professionals. With 11% of the world's population and 24% of the world's illness burden, Sub-Saharan Africa's health issues are recognized as the greatest difficulties. However, the region has just 3% of the world's health workers and receives less than 1% of the global health budget. Additionally, there are too few facilities relative to the densely populated area.

Globally, maternal health is considered one of the healthcare issues, and maternal mortality and morbidity have become issues that require technology to tackle. The majority of avoidable child deaths occur during pregnancy or childbirth. This death has been attributed to preventable complications if women attend regular antenatal care, take routine medication, and receive maternal health education services [1].

These are some of the main explanations why computers (Information Technology (IT)) have been used to assist in various aspects of health care over the years, including diagnosis, treatment, monitoring, and medical records. It is abundantly clear that the majority of these computerized health management systems are deficient in several areas, including the ability of patients and healthcare professionals to engage in real-time prediction of diseases [2].

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 13, Issue 4, pp: (1-11), Month: October - December 2025, Available at: www.researchpublish.com

One of the primary information technologies for antenatal management is the utilization of a machine learning system for healthcare interventions. A machine learning tool is an essential tool for antenatal care and monitoring of pregnant women, and offers services such as support for healthcare providers, pregnant women, and family members to predict and provide adequate healthcare. In addition, it facilitates real-time monitoring and access to critical information, communication, and information dissemination. The machine learning systems implemented in the Internet of Things and mobile technology provide important services, such as reminders for antenatal visits, personalized health tips, and monitoring vital signs, including blood pressure. weight, blood sugar, and temperature [3].

Mobile healthcare monitoring raises the standard of service and lowers expenses in underdeveloped countries, such as Nigeria, and other rural West African environments. One of the major challenges faced by pregnant women is the issue of gestational diabetes during pregnancy [4]. Gestational diabetes mellitus is the glucose intolerance that occurs during pregnancy. It comes with complications such as a higher burden of Cesarean sections, pre-eclampsia, neonatal hypoglycemia, and jaundice in the fetus [5], [6]. Other complications related to the onset of gestational diabetes during pregnancy include spontaneous abortion, secondary infections, fetal malformation, hypertensive disorder, cholestasis, obstructed vaginal delivery, and kidney disease [7].

Gestational diabetes mellitus is the carbohydrate intolerance of a high degree that occurs during pregnancy. Research has shown that gestational diabetes mellitus is a risk factor for other diseases, such as Type II diabetes in later years for women and children developing obesity or Type II diabetes in adulthood [5], [6], [8]. To assess pregnant women for gestational diabetes risk during pregnancy, doctors use various information such as socioeconomic information, medical history, obstetric features, and medical laboratory parameters [9]. The socioeconomic factors include age, educational level, body mass index, and family history of diabetes. Obstetric features such as an oral glucose tolerance test (OGTT), gravida, and parity during the pre-conceptional period provide important information to detect gestational diabetes mellitus. Laboratory parameters such as platelet count, white blood cell count, mean arterial pressure, and glucose in the urine have also provided information features for diagnosing gestational diabetes mellitus [7].

Recently, computational methods have been proposed for predicting to assess pregnant women for the presence of gestational diabetes mellitus. These methods include the statistical and artificial intelligence methods [6]. However, artificial intelligence methods have more advantages over statistical approaches as they identify the relationship between different variables independently. Moreover, they provide highly stable results in various data sizes. Various studies have developed artificial intelligence methods for gestational diabetes mellitus prediction [10], [11], [7], [12], [13]. In their study, Nombo et al [12] developed a multivariate logistic model to predict the likelihood of pregnant women developing gestational diabetes mellitus using family history. Also, Wang et al [6] utilized machine learning algorithms to assess the risk factors that contribute to the onset of gestational diabetes mellitus to determine pregnant women who might be affected during pregnancy. Hu et al [13] implemented extreme gradient boosting algorithms for predicting gestational diabetes mellitus and compared the results with other machine learning models, such as logistic regression. He noted that extreme gradient boosting algorithms produce superior results in terms of accuracy and operating receiver characteristics in predicting the condition. Furthermore, the study utilized demographic characteristics, clinical features, and laboratory parameters. In their recent study, Bigdeli et al [9] explore the use of the extreme gradient boosting algorithm to predict gestational diabetes mellitus in the first trimester of pregnancy and compare the results obtained with other machine learning methods, such as decision tree, multilayer perceptron, k-Nearest Neighbors, Naïve Bayes, and Random Forest algorithms. The comparison indicates the high performance of the extreme gradient boosting algorithm for gestational diabetes mellitus prediction and its potential for real-time implementation for healthcare intervention. The extreme gradient boosting algorithm can model complex systems, provide high prediction accuracy, generalization, interpretability, and versatility for different classification tasks. In addition, it combines different weak models in parallel to create a strong model with reduced classification errors.

Therefore, considering the strengths of the extreme gradient boosting algorithm for different predictive tasks, this study proposes it for gestational diabetes mellitus prediction using data collected from pregnant women in their first trimester.

The remainder of the paper is as follows: Section 2 outlines the methodology adopted for the implementation, Section 3 describes the experimental and hyperparameter settings of the machine learning algorithms, while Section 4 outlines the results and their implications. Section 5 concludes the paper.

II. METHODOLOGY

This section describes the methodology adopted to develop the proposed gestational diabetes-based machine learning model among pregnant women. The section is divided into different subsections and includes the dataset description, preprocessing, feature analysis and normalization, and gestational diabetes prediction model. In addition, the subsection outlined the performance metrics used to evaluate the machine learning model and experimental settings. The detailed flow of the proposed approach is shown in Fig. 1.

2.1. Data Description

The dataset for implementing the gestational diabetes mellitus prediction system was collected from Kaggle (https://www.kaggle.com/datasets/sumathisanthosh/gestational-diabetes-mellitus-gdm-data-set). The dataset contains 3525 instances and fifteen (15) feature sets. It is roughly divided into 2153 women without gestational diabetes mellitus, marked as 0, and 1372 pregnant women with gestational diabetes mellitus, marked as 1. The sample format of the dataset is shown in Fig. 2.

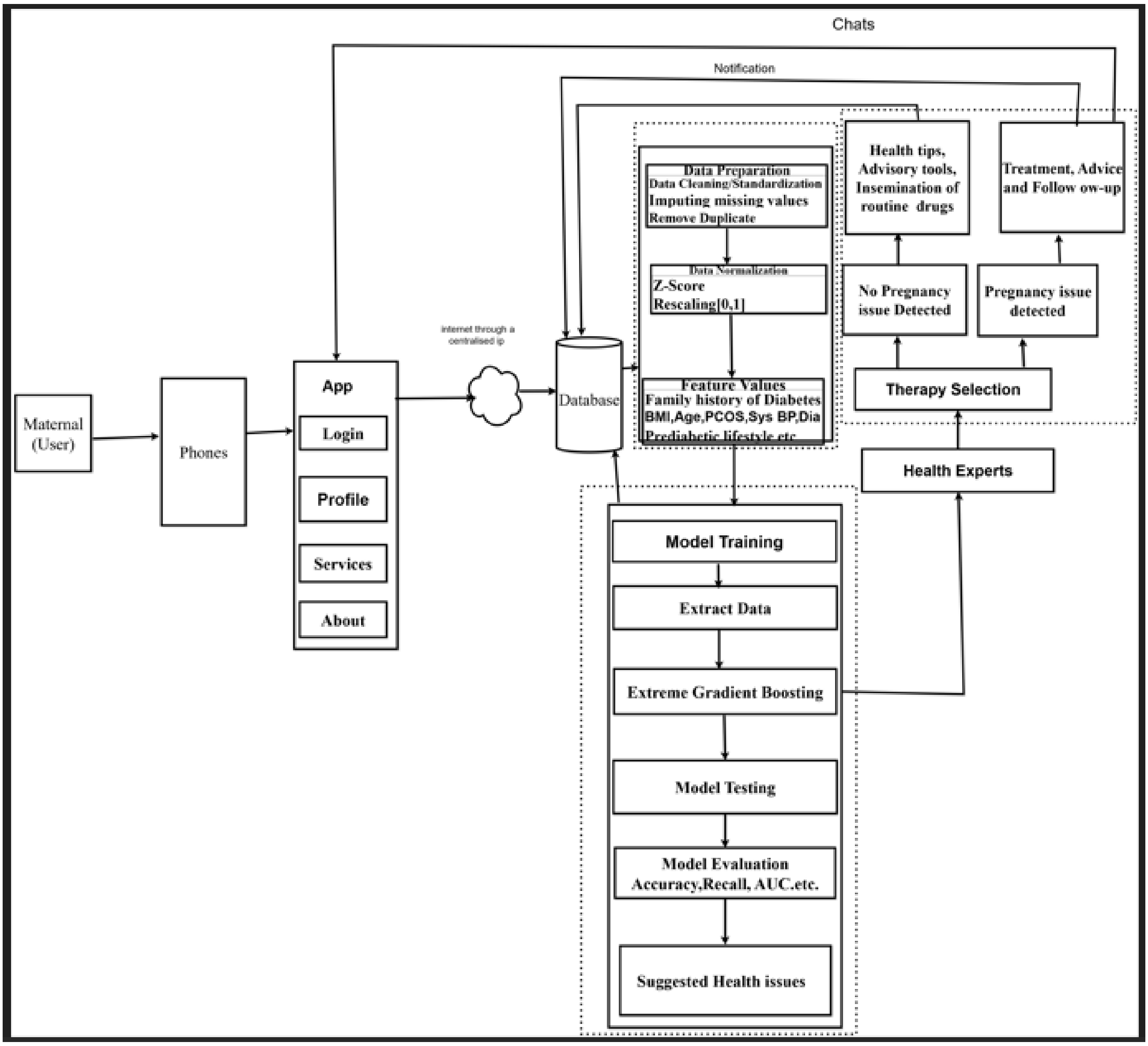

Fig. 1. Proposed gestational diabetes mellitus prediction system

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 13, Issue 4, pp: (1-11), Month: October - December 2025, Available at: www.researchpublish.com

Case Numl	Age	No of Preg	Gestation	BMI	HDL	Family Hist	unexplaine	Large Chilc	PCOS	Sys BP	Dia BP	OGTT	Hemoglob	Sedentary	Prediabete	Class Label(GDM /Non GDM)
1	22	2	1		55	0	0	0	0	102	69		12	0	0	0
2	26	2	1		53	0	0	0	0	101	63		12.4	0	0	0
3	29	1	0		50	0	0	0	0	118	79		14.3	0	0	0
4	28	2	1		51	0	0	0	0	99	70		15	0	0	0
5	21	2	1		52	0	0	0	0	116	65		15	0	0	0
6	29	2	1		51	0	0	0	0	98	63		15.2	0	0	0
7	26	2	1		51	0	0	0	0	94	68		15	0	0	0
8	27	1	0		52	0	0	0	0	116	63		12	0	0	0
9	26	1	0		57	0	0	0	0	108	62		14	0	0	0
10	21	2	1		52	0	0	0	0	98	78		13	0	0	0
11	21	2	1		56	0	0	0	0	100	76		14	0	0	0
12	26	2	1		50	0	0	0	0	110	68		13	0	0	0
13	27	2	1		55	0	0	0	0	105	61		13.6	0	0	0
14	25	2	1		58	0	0	0	0	106	80		15	0	0	0
15	22	1	0		53	0	0	0	0	109	61		15.9	0	0	0
16	22	2	1		57	0	0	0	0	107	80		14	0	0	0
17	27	1	0		57	0	0	0	0	100	61		13	0	0	0
18	21	2	1		58	0	0	0	0	105	78		13	0	0	0
19	28	1	0		50	0	0	0	0	112	77		12.1	0	0	0
20	20	2	1		54	0	0	0	0	97	74		13	0	0	0
21	20	1	0		50	0	0	0	0	93	66		12	0	0	0
22	23	1	0		51	0	0	0	0	116	72		14	0	0	0
23	29	2	1		59	0	0	0	0	94	62		13.9	0	0	0
24	21	1	0		55	0	0	0	0	112	78		14	0	0	0
25	29	2	1		51	0	0	0	0	99	71		15	0	0	0
26	24	1	0		56	0	0	0	0	101	64		13.4	0	0	0
27	21	1	0		55	0	0	0	0	120	68		13.2	0	0	0
28	25	2	1		54	0	0	0	0	104	66		14	0	0	0

Fig. 2. Sample gestational diabetes mellitus dataset.

2.2. Preprocessing

In most cases, the dataset collected for gestational diabetes mellitus prediction is affected by noise, missing values, and duplications. Accordingly, preprocessing methods are used to remove anomalies and duplications in the data. in addition, imputing missing values is performed using different techniques such as the moving average or the variance of each column of data. here, we prepared the data to ensure enhanced machine learning algorithms' performance and data quality. First, the data was converted from *.xls format* to *.csv format*. Second, we performed class labelling to assign a class label to the data. To replace the missing values, a group-based mean was implemented. This means replacing the missing data inside a group or column with the mean values of the non-NaN measurements in a group [14].

2.3. Feature analysis and Normalization

Feature analysis is implemented to identify and outline the most relevant feature attributes that would predict the outcome of gestational diabetes mellitus in pregnant women during antenatal visits. To achieve this, various data and information are collected from pregnant women. Some of the data that are succinctly analysed to determine the onset of gestational diabetes mellitus include age, number of pregnancies, gestational diabetes in previous pregnancies, body mass index, high-density lipoprotein, family history of diabetes, unexplained prenatal loss, large child, or birth defect. Others are Polycystic ovary syndrome, systolic blood pressure, diastolic blood pressure, oral glucose challenge test, haemoglobin, sedentary lifestyle, and prediabetes. The details of the features are highlighted in Table I, while the feature vectors distribution is shown in Fig. 3

TABLE I: ATTRIBUTES/FEATURES OF THE DATASET USED FOR DISEASE PREDICTION

Features	Description	Types	Values
Case Number	Identification number	Integer	[1-3525]
Age	Age	Integer	[20-30]
No of Pregnancy	Number of pregnancies	Integer	[1-2]
Gestation in a previous Pregnancy	Gestational diabetes mellitus during previous pregnancy	Integer	[1, 0]
BMI	Body mass index	float	[126-564]
HDL	High-density lipoprotein	float	[50-60]
Family History	Family history of diabetes	Integer	[0,1]
unexplained prenatal loss	Unexplained prenatal loss	Integer	[0,1]
Large Child or Birth Default	Large child or macrosomia	Integer	[0,1]
PCOS	Polycystic ovary syndrome	Integer	[0,1]
Sys BP	Systolic blood pressure	float	[90-200]
Dia BP	Diastolic blood pressure	Integer	[70-80]

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 13, Issue 4, pp: (1-11), Month: October - December 2025, Available at: www.researchpublish.com

Features	Description	Types	Values
OGTT	Oral glucose tolerance test	float	[0-3]
Hemoglobin	Hemoglobin	float	[9.8-15.0]
Sedentary Lifestyle	Living a sedentary lifestyle	Integer	[0,1]
Prediabetes	Prediabetes before pregnancy	Integer	[0,1]
Target class	Gestational diabetes mellitus predicted	Integer	*No GDM=0; GDM=1*

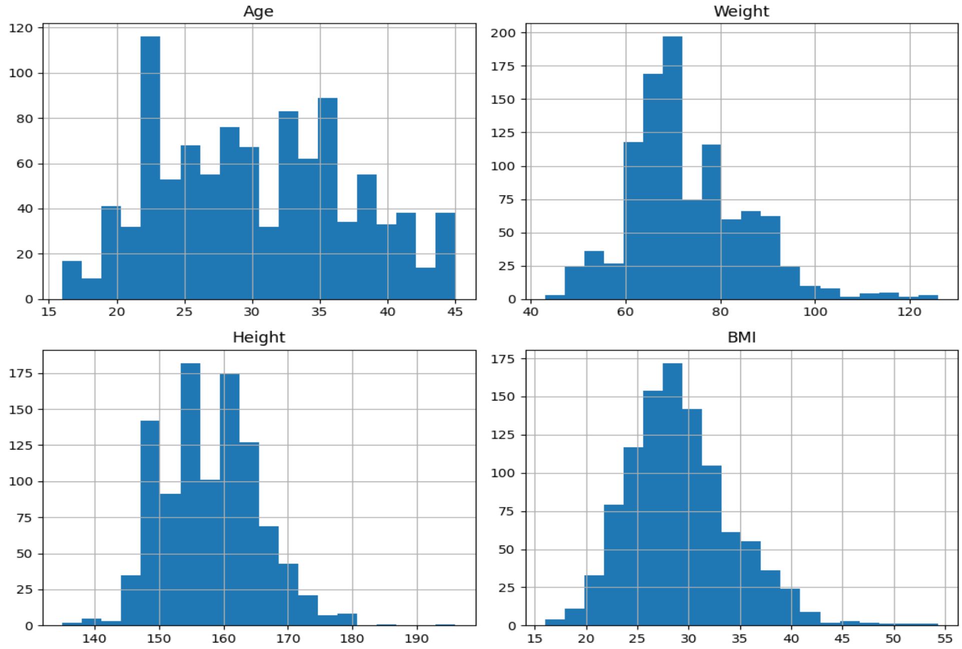

Fig. 3. Feature Distribution

To ensure that the features of the data collected correlated with each other, feature correlation was performed on the data. In this case, data visualization was used to ascertain the discriminative strength of each attribute included in the dataset for implementing the disease prediction system. Here, the heatmap as shown in Fig. 4 was used to visualize the relationship between the data values in the dataset, and the Figure shows the outliers in the feature vector. In the diagram, there are indicators of gestational diabetes mellitus disease symptoms, such as age, number of pregnancies, weight, BMI, etc. In addition, Figure depicts the distributions of the features among the age groups. The features considered include weight, height, and body mass index (BMI).

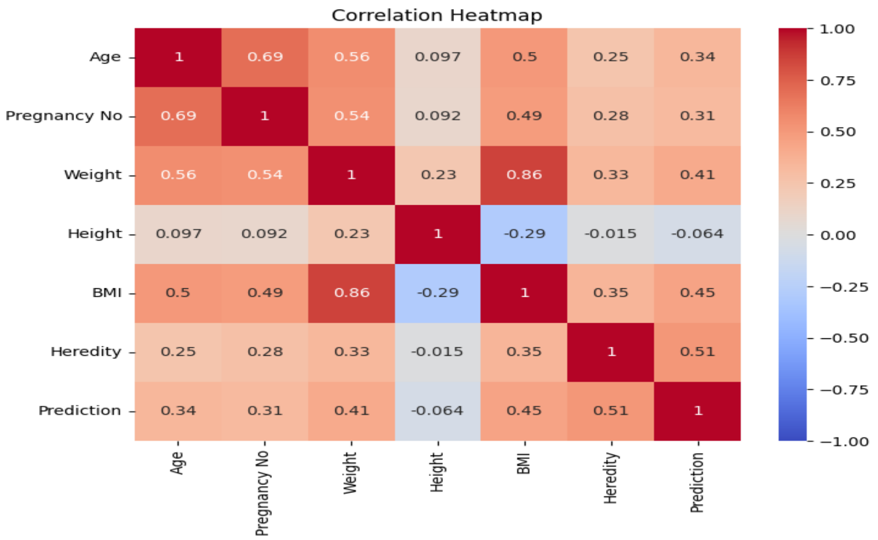

Fig. 4. Correlation Heatmap of the features.

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 13, Issue 4, pp: (1-11), Month: October - December 2025, Available at: www.researchpublish.com

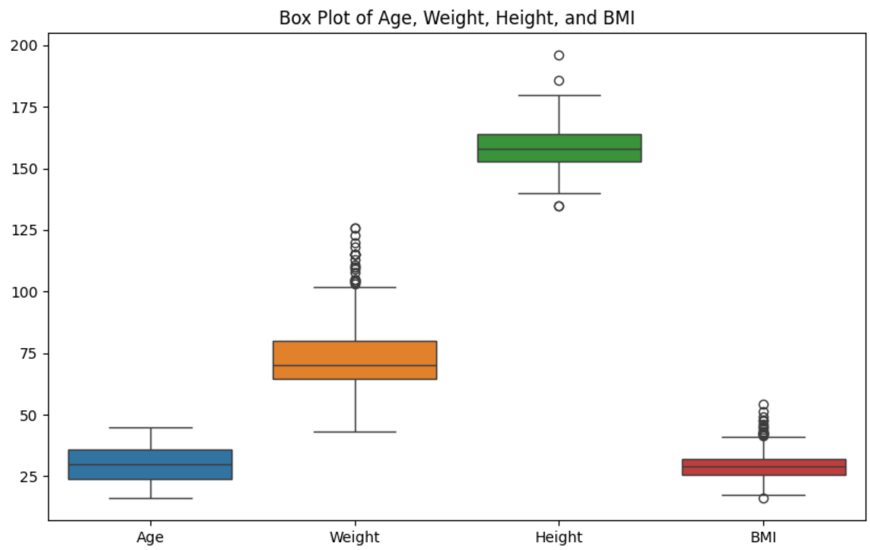

Fig.5 (a). Outlier detection and feature distributions

After identification of features, the features were normalized to zero mean and unit variance to reduce the features to a certain range. Feature normalization is important in a prediction system as it helps to improve its performance. The Z-score normalization was used, where the mean value of the feature vectors was subtracted from the individual feature value point and divided by the standard deviation. Equation 1 below shows the formula for computing Z-score normalization, where $\acute{x}$ represent the computed z-score, $\bar{x}$ is the mean value, x are the individual features and α is the standard deviation.

$$\acute{x} = \frac{\bar{x}-x}{\alpha} \qquad (1)$$

The normalized features were combined into a master feature vector and saved as .csv for the gestational diabetes mellitus prediction system implementation. In addition, some of the feature cluster distributions such as such as weight, height, and age, are shown in Fig. 5(b).

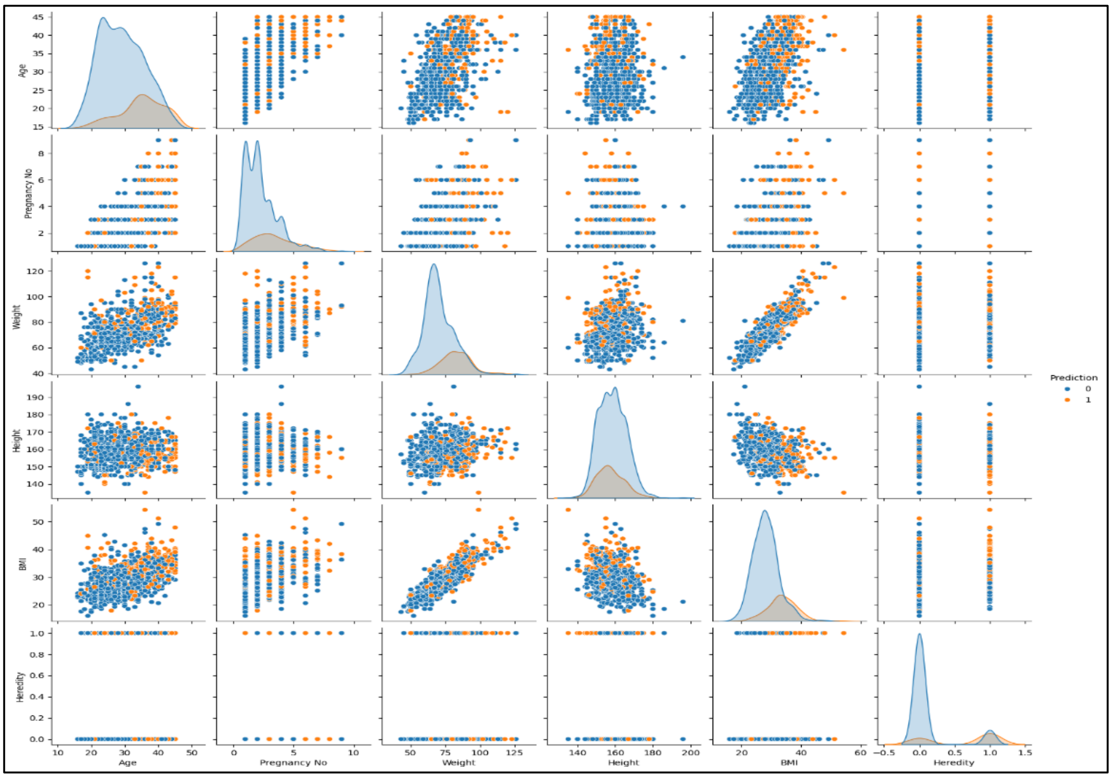

Fig. 5 (b). Feature importance distribution

2.4. Gestational Diabetes Mellitus Prediction System

The thesis utilised supervised machine learning models to build the gestational diabetes mellitus disease prediction system. Supervised machine learning is the type of machine learning where labelled datasets are used to train the prediction models. Here, we proposed three machine learning models: Extreme Gradient Boosting (XGBoost), Random Forest (RF), and Support Vector Machine (SVM) to build the system.

- *XGBoost* is an efficient and accurate ensemble learning algorithm that utilizes decision tree models as base learners and uses an average voting method to produce a final prediction.
- *Random forest* model is also an ensemble learning model that uses a decision tree model as a base learner. It creates a set of decision trees from a subset of prediction features and uses a bootstrapping approach to produce the final prediction.
- *A support vector machine* is a machine learning algorithm that implements a kernel with maximum margin, which separates the data into classes to distinguish between features that might result in pregnancy complications and those that wouldn't result in pregnancy complications. Nevertheless, the Support Vector Machine model is memory-intensive and slow when implemented with a large volume of data collected from pregnant women.

The three machine learning models were implemented to ensure comprehensive prediction of diseases during antenatal care, decision-making by healthcare providers, and to minimize the onset of gestational diabetes mellitus-based complications.

2.5. Performance Evaluation Metrics

To evaluate the performance of the proposed gestational diabetes mellitus prediction system using the collected data, different performance metrics were utilized. These performance metrics include accuracy, recall, precision, F-measures, Area under the curve (AUC), and confusion matrix. Accuracy measures the ratio of correctly predicted gestational diabetes mellitus to the total number of both presence and absence of gestational diabetes mellitus-related disease among pregnant women who visit the clinic/hospital. Area under the curve (AUC) measures the performance of the disease prediction system, in which the system can predict cases of women with gestational diabetes mellitus more accurately than women with no gestational diabetes mellitus. The AUC shows the performance of the classification model using two parameters (True Positive Rate and False Positive Rate) at various threshold values. AUC ranges in value from 0 to 1. Then, a confusion matrix combines two or more evaluation or classification models for which set of test data whose true values are known. Furthermore, these performance metrics are mathematically represented as:

Accuracy (ACC) $= \frac{(TP+TN)}{(TP+TN+FP+FN)};$

Recall $= \frac{TP}{(TP+FN)};$

Precision $= \frac{TP}{(TP+FP)};$

Where TP, TN, FP, and FN represent true positive, true negative, false positive, and false negative, respectively.

III. EXPERIMENTAL SETTINGS

The implementation of the different phases of the system, such as preprocessing, feature analysis, and algorithm creation, was done on a computer system running the Windows 11 operating system. The gestational diabetes mellitus disease prediction system was implemented in Python. With a Random-Access Memory (RAM) installed capacity of 8GB, the computer system uses an Intel Core i7 processor running at 3.400 GHz. Parts for training and parts for testing were taken from the entire pre-processed data set. Here, 70% of the data was used for training the disease prediction system, while 30% was utilized to test the developed system. Using the train-test data partitioning approach ensures uniformity and reduces complexity. Important hyperparameter settings for both XGBoost, Random Forest, and Support Vector Machine algorithms are outlined in Table II. These parameter settings are the defaults and were selected based on the results of an empirical evaluation of machine learning algorithms for the prediction of cardiac disease [15]. Additionally, applying some of the machine learning models' default settings would guarantee the algorithms' reproducibility in similar settings.

TABLE II. PARAMETER VALUES FOR EACH MACHINE LEARNING MODEL

Machine learning algorithms	Parameter Tuning
XGBoost	Booster=gbtree;silent=0;nthread=50;gamma=0;max_depth=6; subsample=1
Random forest (RF)	n_estimator=100;criterion=gini;max_depth=None;min_sample_split=2;random_state=None;verbose=0
Support vector machine (SVM)	Kernel=rbf;C=1;gamma=scale;degree=3;max_iter=-1;random_state=None;verbose=false

IV. RESULTS AND DISCUSSION

In this section, the results obtained with the implementation of the proposed machine learning-based disease prediction system are presented. The biodata information collected during patient registration, visitation, and similar data crawled online was saved in the database and utilized to develop the disease prediction system. First, the data were saved in comma-separated values (CSV) alongside the target labels and sent to the listed machine learning algorithms for gestational diabetes mellitus disease prediction. The results obtained are presented in Table III. From the table, the Support Vector Machine achieved a prediction accuracy of 83.44%, followed by the XGBoost algorithm achieved a prediction accuracy of 83.11%, and Random Forest, which obtained an accuracy of 82.79%. In terms of performance accuracy, the algorithms have similar performance with only marginal differences, with the Random Forest algorithm.

TABLE III. PERFORMANCE RESULTS OF THE GESTATIONAL DIABETES DISEASE PREDICTION SYSTEM

Methods	Accuracy	Precision	Recall	F1-Score	AUC
XGBoost Algorithm (XGB)	83.11%	0.7121	**0.5603**	**0.6234**	0.8141
Random Forest Algorithm (RF)	82.79%	0.7080	0.5356	0.6055	**0.8414**
Support Vector Machine (SVM)	**83.44%**	**0.7401**	0.4623	0.5731	0.7822

The performance results and ROC curves of these algorithms are presented in Figs. 6 and 7, respectively.

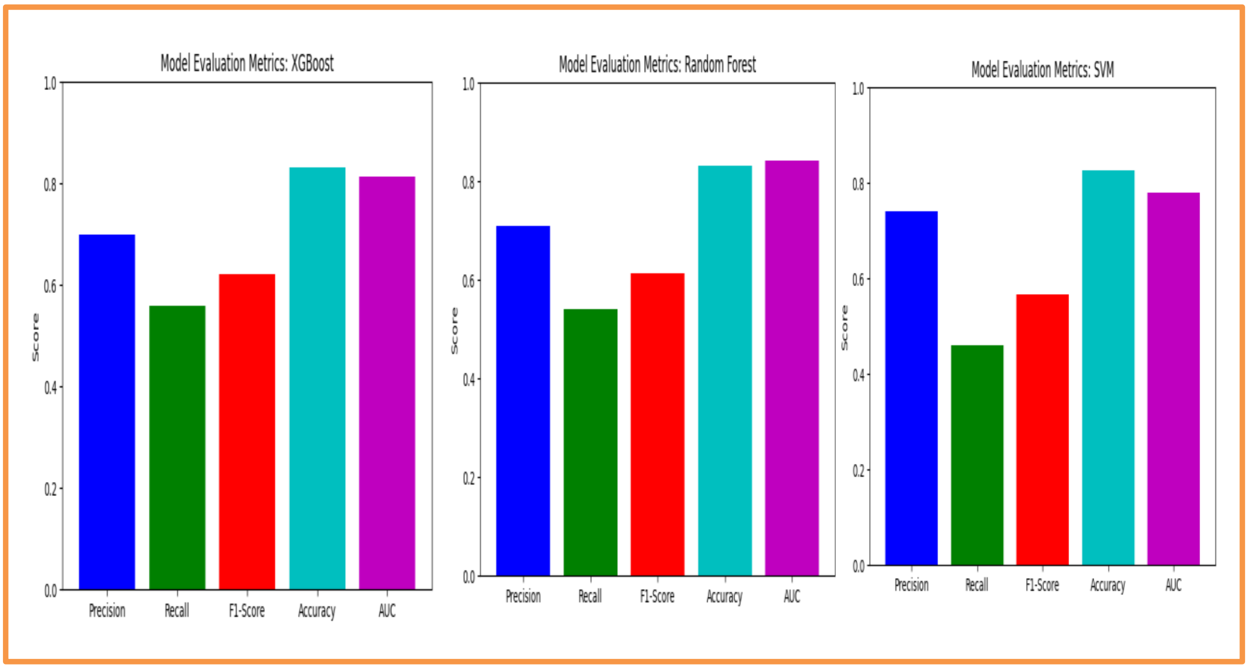

Fig. 6. Performance results obtained using the three algorithms

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 13, Issue 4, pp: (1-11), Month: October - December 2025, Available at: www.researchpublish.com

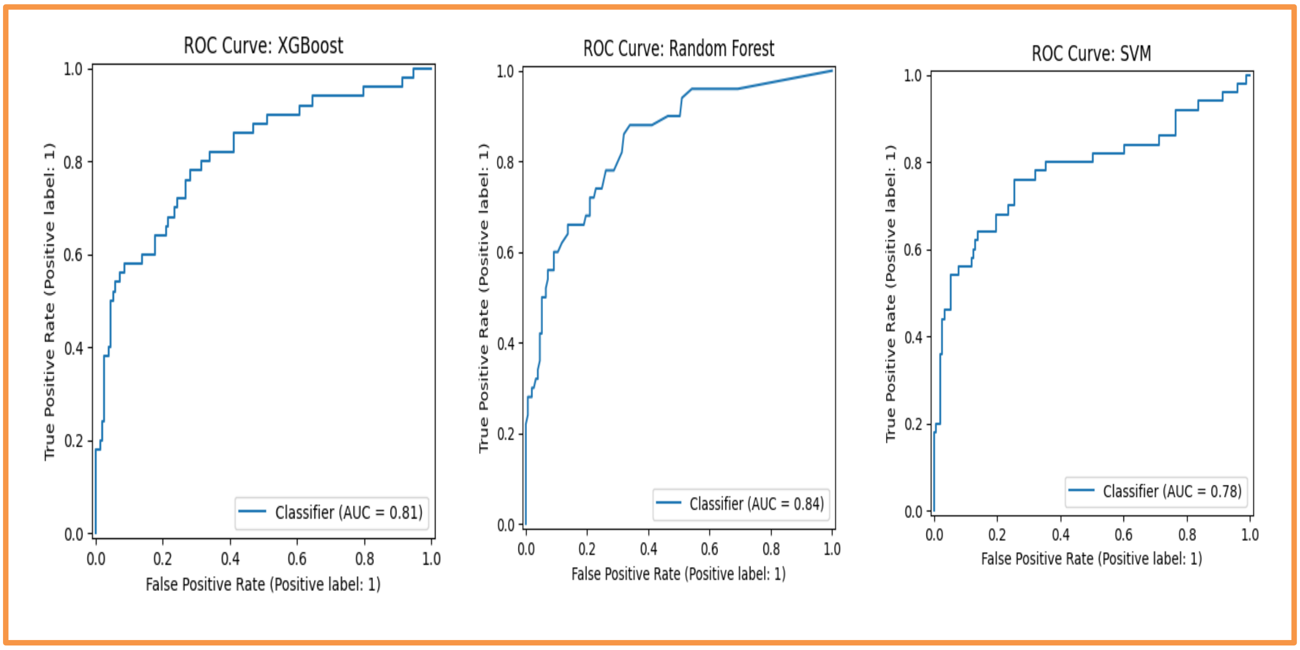

Fig. 7. ROC Curve of the three prediction algorithms

The confusion matrices of the proposed gestational diabetes prediction system using XGBoost, Random Forest, and Support Vector Machine are shown in Fig.8, Fig. 9, and Fig. 10, respectively. A confusion matrix helps to visualize and represent the prediction outcome of the developed system. It represents the true positive (actual number of gestational diabetes mellitus cases) and false positive (actual number of gestational diabetes mellitus cases incorrectly predicted) using the machine learning algorithms. Here, “1” in the figures represents the presence of gestational diabetes mellitus condition while “0” represents the absence of gestational diabetes mellitus in pregnant women.

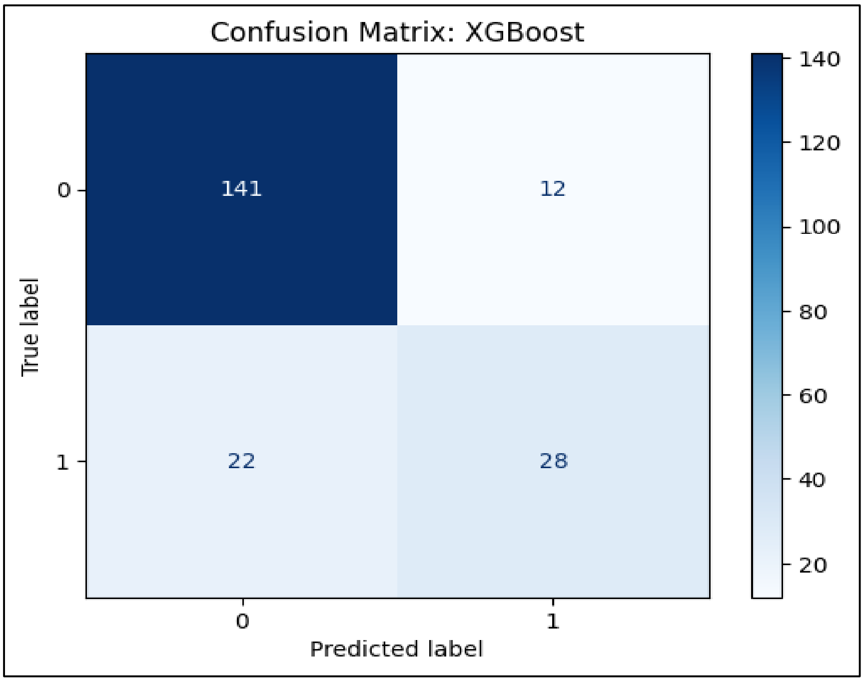

Fig. 8. Confusion matrix of the XGBoost algorithm

Fig. 8 represents the confusion matrix of the implemented disease prediction systems. From the figure, the XGBoost algorithm predicted 28 cases of pregnant women with gestational diabetes mellitus at the health facility, and 141 pregnant women with no cases of gestational diabetes mellitus. However, the system wrongly predicted 22 women to have diabetes and 12 pregnant women with no gestational diabetes, even though there were traces of the disease in them.

From Fig. 9, the Random Forest algorithm predicted 27 cases of pregnant women with the condition, while 142 cases of pregnant women with the health facility had no trace of gestational diabetes mellitus. Although gestational diabetes might not be very common within the facility, it poses significant maternal health challenges for women with the condition [14], [16].

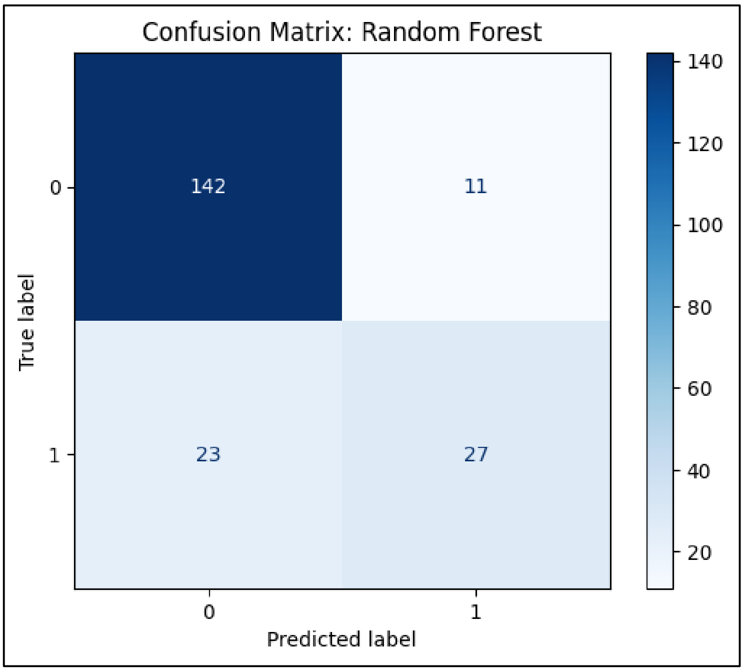

Fig. 9. Confusion matrix of the Random Forest algorithm

Similar results were obtained using Support Vector Machine algorithms, as shown in Fig. 10.

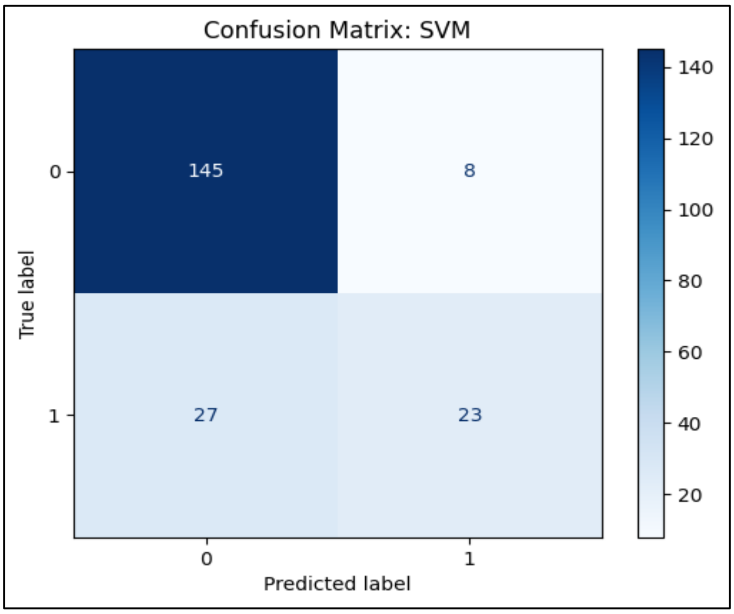

Fig. 10. Confusion matrix of the Support Vector Machine algorithm

V. CONCLUSION

This study implemented machine learning methods for gestational diabetes prediction using a publicly available dataset. Specifically, the study implemented extreme gradient boosting models, random forest, and support vector machine models to assess important features that contribute to the onset of GDM in women in their first trimester. The implementation of the models utilized important features such as age, body mass index, family history, blood pressure, sedentary lifestyle, etc., after preprocessing and data normalization. The algorithms achieved an average accuracy of 83.11% and an average Fmeasure of 0.4702. The results indicate the effectiveness of the models to predict gestational diabetes mellitus and other health-related issues during pregnancy. Future studies will focus on the implementation of deep learning methods for gestational diabetes mellitus and the explainability of the models.

REFERENCES

[1] D. Ahmad, I. Mohanty, A. Hazra, and T. Niyonsenga, "The knowledge of danger signs of obstetric complications among women in rural India: evaluating an integrated microfinance and health literacy program". *BMC pregnancy and childbirth*, *21*, 1-23, 2021.

[2] I. Maimunatu, M. Kabiru, Y. Yakubu, and A. Aminu, "A Proposed Secured Health Monitoring System for the Elderly using Blockchain Technology in Nigeria". *Journal of Electronics, Computer Networking and Applied Mathematics ISSN: 2799-1156 Vol: 02, No. 02, Feb-Mar 2022 http://journal.hmjournals.com/index.php/JECNAM*

[3] S. Soni, K. A. Sharma, K. Yadav, S. Singhal, and R. Kaur, "Bridging gaps in antenatal care: Implementation of the WHO ANC model using a web-based mobile application at different levels of the healthcare system". *International Journal of Medical Informatics*, *180*, 105277, 2023.

[4] T. A. Azeez, T. Abo-Briggs, and A. S. Adeyanju, "A systematic review and meta-analysis of the prevalence and determinants of gestational diabetes mellitus in Nigeria". *Indian journal of endocrinology and metabolism*, *25*(3), 182-190, 2021.

[5] R. AlSaad, A. Elhenidy, A. Tabassum, N. Odeh, E. AboArqoub, A. Odeh, and J. Sheikh, "Artificial Intelligence in Gestational Diabetes Care: A Systematic Review". *Journal of Diabetes Science and Technology*, 1932296825135 5967, 2025.

[6] X. Wang, Y. Wang, S. Zhang, L. Yao, and S. Xu, "Analysis and prediction of gestational diabetes mellitus by the ensemble learning method". *International Journal of Computational Intelligence Systems*, *15*(1), 72, 2022.

[7] Y. Kaya, Z. Bütün, O. Çelik, E. A. Salik, T. Tahta, and A. A. Yavuz, "The early prediction of gestational diabetes mellitus by machine learning models". *BMC Pregnancy and Childbirth*, *24*(1), 574, 2024.

[8] H. X. Zhang, H. Guo, and J. X. Wang, "Research on type 2 diabetes mellitus has developed precise prediction models based on the XGBoost algorithm". *Chin J Lab Diagn*, *22*(3), 408-12, 2018.

[9] S. K. Bigdeli, M. Ghazisaedi, S. M. Ayyoubzadeh, S. Hantoushzadeh, and M. Ahmadi, "Predicting Gestational Diabetes Mellitus in the first trimester using machine learning algorithms: a cross-sectional study at a hospital fertility health center in Iran". *BMC Medical Informatics and Decision Making*, *25*(1), 3, 2025.

[10] B. Wu, H. Huang, Q. Yao, R. Deng, H. Li, and T. Liang, "Application of big data and artificial intelligence methods in the prediction of gestational diabetes mellitus". *Chin J Health Inf Manag*, *14*(6), 832-837, 2017.

[11] S. Kumari, D. Kumar, and M. Mittal, "An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier". *International Journal of Cognitive Computing in Engineering*, *2*, 40-46, 2021.

[12] A. P. Nombo, A. W. Mwanri, E. M. Brouwer-Brolsma, K. L. Ramaiya, and E. J. Feskens, "Gestational diabetes mellitus risk score: a practical tool to predict gestational diabetes mellitus risk in Tanzania". *Diabetes research and clinical practice*, *145*, 130-137, 2018.

[13] X. Hu, X. Hu, Y. Yu, & J. Wang, "Prediction model for gestational diabetes mellitus using the XGBoost machine learning algorithm". *Frontiers in Endocrinology*, *14*, 1105062, 2023.

[14] A. Sumathi and S. Meganathan, "Ensemble Classifier Technique to Predict Gestational Diabetes Mellitus (GDM)". *Computer Systems Science & Engineering*, *40*(1), 2022.

[15] C. Ayón, "Unpacking immigrant health: Policy, stress, and demographics. *Race and Social Problems*, *10*, 171-173, 2018.

[16] B. Viola and T. Jones, "The algorithm for recognizing and detecting health issues". *Proceedings of the 23rd international conference on Machine learning. ACM, 233–240, 2020.*