Evaluation of A Novel Risk Factor Screening Tool for Gestational Diabetes Mellitus: A Machine Learning Based Predictive Method

Usha Adiga[1,*], Tirthal Rai[2], Sampara Vasishta[1], Banubadi Anil Kishore[1], Tunuguntla Amulya[1], Lokesh Ravi[3]

[1]*Department of Biochemistry, Apollo Institute of Medical Sciences and Research Chittoor, Murukambattu - 517127, Chittoor, Andhra Pradesh, India*

[2]*Department of Biochemistry, A.J. Institute of Medical Sciences and Research Centre, Mangalore - 575004, Karnataka, India*

[3]*Centre for Digital Health and Precision Medicine, The Apollo University, Chittoor, Andhra Pradesh, 517127, India*

Abstract: *Introduction*: Gestational diabetes mellitus (GDM) is a significant pregnancy complication linked to adverse outcomes for both mother and child. Early identification of high-risk individuals is crucial for effective management and prevention for the onset/progression of the GDM. Our study aims to a) evaluate the effectiveness of a newly developed machine learning based risk factor screening tool for predicting GDM and b) to compare its predictive performance against established models and current literature.

Methods: This study explored SNP data from the leptin (*LEP*) and leptin receptor (LEPR) genes to develop machine learning models for predicting gestational diabetes mellitus (GDM). It included data preprocessing, such as cleaning and feature selection, focusing on genetic markers, metabolic parameters, and demographic information. Various algorithms, including Logistic Regression, Decision Trees, and Random Forests, were used, and their performance was evaluated using metrics like accuracy and ROC-AUC to determine the best model for GDM prediction.

Results: The newly developed screening tool demonstrated a sensitivity of 85%, specificity of 78%, positive predictive value (PPV) of 68%, and negative predictive value (NPV) of 90% in predicting GDM. Comparatively, machine learning models showed higher sensitivity (90-95%) but lower specificity (65-75%).

Conclusion: The developed risk factor screening tool is a viable method for predicting GDM, with accuracy metrics comparable to advanced machine learning models and established literature. Future research should focus on refining these tools and exploring their integration into routine prenatal care to enhance early detection and intervention strategies for GDM.

Keywords: Artificial Intelligence, ROC, Naïve Bayes, XG Boost.

1. INTRODUCTION

Gestational diabetes mellitus (GDM) manifests during pregnancy, posing significant health risks to both the mother and the foetus [1]. This is characterized by high blood glucose levels and can predispose to complications such as preeclampsia, macrosomia and increased rates of caesarean delivery. Moreover, women who experience GDM are at a higher risk of developing type 2 diabetes (T2D) over a period of time [2]. Despite having a comparatively lower body mass index (BMI), Asian women are prone to develop GDM when compared with other ethnic groups [3]. Approximately 10.1% - 20% of people in Eastern and Southeastern Asian communities were found to have GDM. A recent study reported a GDM prevalence of 9% in India [4]. Considering the prevalence and potential complications associated with GDM, early detection and intervention can greatly improve maternal and foetal health outcomes. Nevertheless, opinions about the most effective GDM screening technique are divided [5]. Traditional screening methods, such as the oral glucose tolerance test, can be cumbersome and may not always be predictive of GDM. Thus, there is a growing interest in utilizing biomarkers and advanced machine learning techniques to improve the predictability of GDM [6]. Therefore, in the present study, we aimed to develop robust machine learning models capable of predicting gestational diabetes mellitus using specific biomarkers, including the genotype of leptin and leptin receptor (LepR) levels which regulate the energy homeostasis, along with other clinical features. Further, we compared the algorithms in terms of accuracy, precision, recall and F1-score which are the performance metrics associated with machine learning. By leveraging the predictive power of machine learning, we aim to enhance the early detection of GDM, timely interventions and better management of the pathological condition.

*Address correspondence to this author at the Department of Biochemistry, Apollo Institute of Medical Sciences and Research Chittoor, Murukambattu - 517127, Chittoor, Andhra Pradesh, India; Tel: 8277781638;
E-mail: ushachidu@aimsrchittoor.edu.in

E-ISSN: 1929-6029/25

2. METHODOLOGY

2.1. Data Source

This study analysed SNP data from the leptin gene (*LEP*) and leptin receptor gene (*LEPR*) in relation to GDM. The primary objectives are to develop machine learning algorithms for predicting GDM and to compare their performance against established models.

We utilized the SNP data from the Biostudies database (Accession number: S-BSST854), which includes genotyping and metabolic parameters [7]. Specifically, focus was on the polymorphisms *LEP* (rs7799039) and *LEPR* (rs1137101) and their association with insulin resistance in pregnant women.

Study Setting: The study was conducted in Central Research Laboratory of K.S.Hegde Medical academy and Department of OBG, K.S.Hegde Charitable Hospital of Nitte University, Mangaluru, Karnataka, India. This was a collaborative study which included Department of Biochemistry, OBG and Pharmacology.

Study Subjects: GDM subjects (n=100) diagnosed based on 75 gm oral GTT (OGTT) as per ADA 2016 criteria were taken as cases. Hundred gestational age and BMI matched normal glucose tolerant pregnant women were considered as control group. Ethical consent was obtained from the study participants.

Exclusion Criteria: Subjects with multiple pregnancies, known pre-gestational diabetes, pregnancies complicated by major fetal malformations or known major cardiac, renal or hepatic disorders, pregnancy-induced hypertension were excluded.

Type of Study: Prospective Cross sectional**.**

Sample Size Calculation: There are not many such studies in the literature so far estimating the correlation of gene and gene polymorphisms in the Indian population. However, by extracting the information derived from the studies published so far on LEP G2548A and considering the prevalence of GDM to be 9.0%, we would require a sample size of 131 patients to design a study with 4% absolute precision and 95% confidence. Due to financial constraints, we restrict the sample size to 100 each for cases and control.

Parameter	Normal Glucose Tolerant Pregnant Women	GDM subjects
Mean Age (years)	27.08 ± 3.73	29.62 ± 4.3
Mean BMI (kg/m²)	25.86 ± 5.86	25.78 ± 6.84
Gestational Age	26.1 ± 1.54	25.87 ± 1.21

Data Collection

The dataset utilized in this study includes a comprehensive set of features relevant to gestational diabetes. These features include age, sex, genotype of leptin, C peptide, cholesterol, fasting blood sugar, leptin receptor (LepR) levels, serum leptin levels, HDL, serum insulin, LDL, triglycerides (TG), and the target variable, GDM status. The data is gathered from clinical records and laboratory tests, ensuring a robust dataset for analysis.

Data Preprocessing

Data preprocessing is a crucial step to ensure the quality and reliability of the machine learning models. This involves:

- **Data Cleaning**: Handling missing values, removing outliers, and ensuring data consistency.
- **Feature Selection**: Selecting relevant features for predicting GDM, with a primary focus on the genotype of leptin and LepR levels.
- **Data Standardization**: Standardizing the features to have a mean of 0 and a standard deviation of 1, ensuring that all features contribute equally to the model.

Model Selection

Several machine learning algorithms were selected to predict GDM, each with unique strengths:

1. **Logistic Regression**: A baseline linear model for binary classification, useful for understanding the relationship between predictors and GDM.
2. **Decision Tree**: Captures non-linear relationships and provides an interpretable model.
3. **Random Forest**: An ensemble method that reduces overfitting and improves predictive performance through multiple decision trees.
4. **Gradient Boosting (XGBoost)**: Known for handling complex data structures and achieving high predictive accuracy.
5. **Support Vector Machine (SVM)**: Effective in high-dimensional spaces and for non-linear data.

6. **K-Nearest Neighbors (KNN)**: Simple and effective for smaller datasets, capturing local patterns.

7. **Naive Bayes**: Efficient and simple, especially if features are independent.

8. **Neural Network (MLPClassifier)**: Models complex relationships through multiple layers of neurons.

9. **AdaBoost**: An ensemble method that boosts the performance of weak learners by focusing on difficult cases.

Model Training and Evaluation

To evaluate the performance of each model:

- **Train-Test Split**: The data is split into training and testing sets.
- **Model Training**: Each model is trained using the training set.
- **Performance Metrics**: Models are evaluated using accuracy, precision, recall, F1-score, and ROC-AUC score.
- **Comparison of Algorithms**: The performance of all models is compared to identify the best-performing algorithm.

ROC Curve Analysis

ROC (Receiver Operating Characteristic) curves were plotted for each algorithm to visualize their performance. The Area Under the Curve (AUC) scores are compared to determine the most effective model for predicting GDM. This analysis helps in understanding the trade-offs between sensitivity and specificity for each model.

Implementation and Interpretation

After identifying the best-performing model, the feature importance scores were analyzed to determine the most significant predictors of GDM. The final model is interpreted and validated, ensuring it is suitable for clinical use. The goal is to provide an interpretable and reliable model that healthcare professionals can use for early prediction and management of GDM.

Next, feature selection was undertaken to identify key features for the machine learning models. This process focused on three main categories: SNP genotypes, specifically *LEP* and *LEPR*; metabolic parameters, including glucose and insulin levels; and demographic data such as age and BMI. The dataset utilized in this study includes a comprehensive set of features relevant to gestational diabetes such as age, sex, genotype of leptin, leptin receptor (LepR), C peptide, total cholesterol (TC), fasting blood sugar, serum leptin levels, high-density lipoprotein cholesterol (HDLc), serum insulin, Low-Density Lipoprotein cholesterol (LDLc), triglycerides (TG) and the GDM status. The data is gathered from clinical records and laboratory tests, ensuring a robust dataset for analysis. Feature selection focused on identifying relevant predictors, emphasizing genetic markers such as leptin and LepR levels known to influence GDM risk. Subsequently, data standardization normalized features to a mean of 0 and a standard deviation of 1, enhancing model performance by equalizing the impact of all variables and minimizing scale-related biases. Finally, normalization was performed to standardize the data, ensuring uniformity across the features, particularly for the metabolic parameters, which helps in improving the model's performance. These preprocessing steps were crucial in optimizing data quality, preparing it effectively for machine learning algorithms, and improving the accuracy and interpretability of GDM predictions based on genetic and clinical data.

2.2. Machine Learning Based Approaches

In selecting machine learning algorithms, each was chosen based on its unique strengths tailored to different aspects of the data. Logistic Regression provided insights into linear relationships and probabilities, while Decision Trees captured non-linear interactions in an interpretable manner. Random Forests and Gradient Boosting methods addressed overfitting and complex data structures through ensemble learning techniques, enhancing predictive accuracy. Support Vector Machines (SVM) excelled in high-dimensional spaces and non-linear classifications, while K-Nearest Neighbors (KNN) and Naive Bayes offered simplicity and efficiency, respectively. Neural networks model intricate relationships through layers of neurons, while AdaBoost combines weak learners to enhance predictive performance. During model training and evaluation, the dataset was split into training and testing sets for training each algorithm. Performance metrics such as accuracy, precision, recall, F1-score, and Receiver Operating Characteristic (ROC) -Area Under the Curve (AUC) score were computed to compare and identify the most effective algorithm for GDM prediction. ROC curve analysis visualized each

model's trade-offs between sensitivity and specificity, informing the selection of the best-performing model. ROC (Receiver Operating Characteristic) curve was plotted using Python programme to understand the performance of various learning models.

3. RESULTS

3.1. Various Machine Learning Models Exhibit Different Predictive Capabilities in Distinguishing between GDM and Non-GDM Cases

The logistic regression model's results indicate the following: For class 0 (non-GDM), the precision is 72%, meaning 72% of the predicted non-GDM instances are correct, and the recall is 90%, meaning 90% of actual non-GDM instances were correctly identified, leading to an F1-score of 80%. For class 1 (GDM), the precision is 82%, but the recall is lower at 56%, indicating a higher number of false negatives, resulting in an F1-score of 67%. The model's overall accuracy is 75%, suggesting it correctly classifies 75% of instances (Figure **1A**). The macro average metrics (precision, recall, F1-score) are around 73-77%, treating all classes equally. The weighted averages, considering class support, revealed a precision of 76%, recall of 75% and F1-score of 74%, indicating the model performs reasonably well, particularly given the imbalanced class distribution (Figure **1B**).

The Decision tree Model results indicate that in non-GDM, the precision is 71% and the recall is 75%. These results showed F1-score of 73%, reflecting a balanced trade-off between precision and recall. For GDM (class 1), the precision is 67%, while the recall is 62%, meaning only 62% of actual GDM cases were correctly identified. This leads to an F1-score of 65%, showing a moderate balance between precision and recall but highlighting a higher rate of false negatives (Figure **1C**). Overall, the model has an accuracy of 69% and the macro average metrics (precision, recall, F1-score) are all 69%, treating all classes equally. Similarly, the weighted averages, which consider the support of each class, are also 69%, indicating consistent performance across the dataset (Figure **1D**). Given these results, the model performs reasonably but might benefit from exploring other algorithms such as Random Forest, Gradient Boosting, or neural networks. Additionally, improving the balance between precision and recall, especially for GDM, is crucial to reduce the number of missed GDM cases (false negatives), which could have significant clinical implications.

The random forest model's classification report reveals a strong performance for non-GDM with a precision of 60%, recall of 93%, and F1-score of 73%, indicating good identification of non-GDM cases. However, for GDM, the model shows a precision of 71% but a much lower recall of 21%, leading to an F1-score of 32% (Figure **1E**). Overall, the model's accuracy is 61%, with a macro average precision of 66%, recall of 57%, and F1-score of 52%, suggesting moderate overall performance but a bias towards the majority class (non-GDM) (Figure **1F**).

The gradient boost model's classification report shows an overall accuracy of 67% and precision for non GDM is 65%, and for GDM, it is 70%. Recall for non GDM is 85%, while for GDM it is 44%, showing the model is better at identifying non-gestational diabetes cases. The F1-score is 0.74 for class 0 and 0.54 for class 1 (Figure **1G** and **1H**).

Support vector machine classifier's performance reveals an overall accuracy of 56%. Precision for non GDM is 56%, with a perfect recall of 100%, indicating the model correctly identifies all non-gestational diabetes cases but has some false positives. However, the model fails to predict any instances of gestational diabetes, with both precision and recall at 0%, leading to an F1-score of 0.71 for class 0 and 0.00 for class 1 (Figure **1I** and **1J**).

K-Nearest Neighbors (KNN)shows an overall accuracy of 58% and for class 0, the precision is 58% and recall is 95%, indicating the model is good at identifying non-gestational diabetes cases but has some false positives. For class 1 (gestational diabetes), the precision is 67%, but the recall is only 12%, showing the model struggles to correctly identify gestational diabetes cases, leading to many false negatives. The F1-score is 0.72 for class 0, reflecting a good balance between precision and recall, but only 0.21 for class 1, indicating poor performance in predicting gestational diabetes (Figure **1K** and **1L**).

Neural network classifier shows accuracy of 58%, precision for class 0 is 61%, while for gestational diabetes, it's 54%. Recall for class 0 is 70%, and for class 1, it's 44%, showing better performance in identifying non-gestational diabetes cases compared to gestational diabetes cases. The F1-scores are 0.65 for class 0 and 0.48 for class 1, reflecting a balanced performance for non-gestational diabetes but moderate performance for gestational diabetes (Figure **1M** and **1N**).

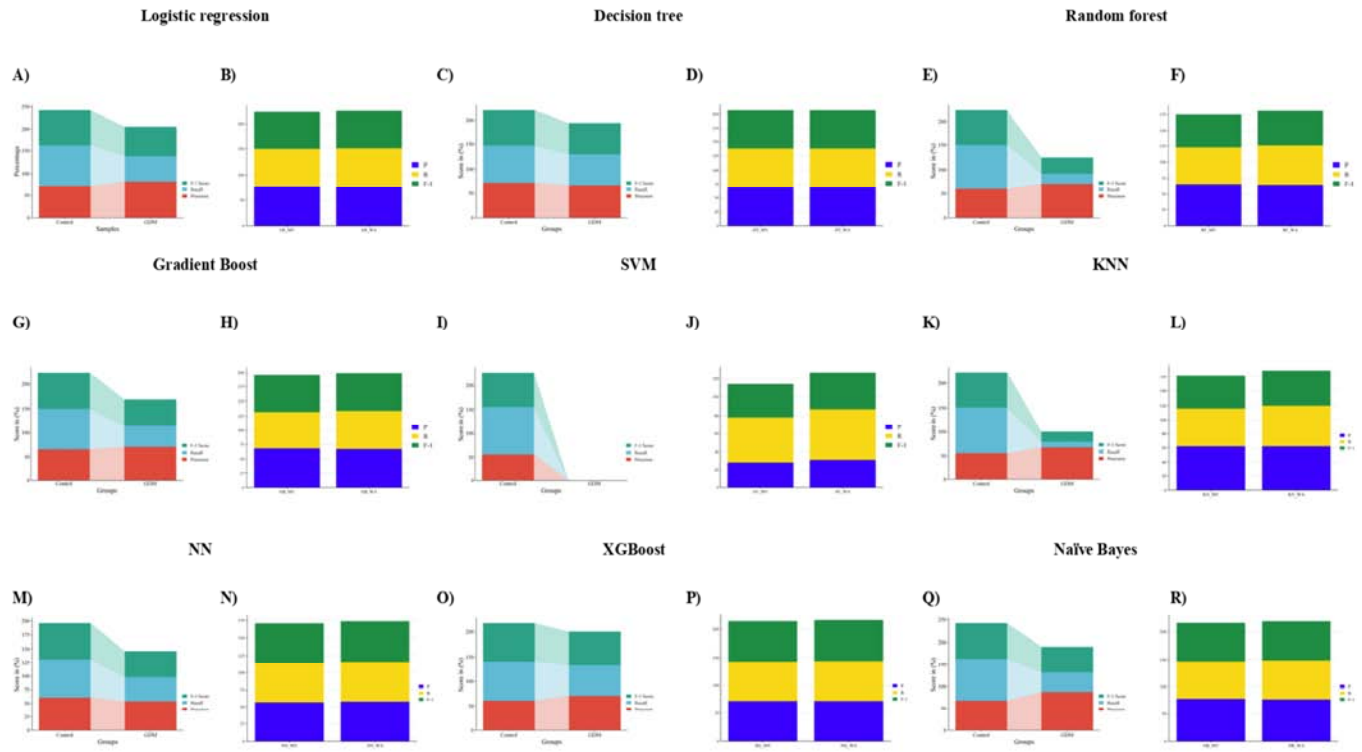

Figure 1: The graph represents precision, Recall and F-1 scores, macro average and weighted average of various ML models in terms of percentage.

XG boost model demonstrates a 72.22% accuracy, precision is 73% for no gestational diabetes and 71% for gestational diabetes. Recall is 80% for class 0 and 62% for class 1, F1-scores of 0.76 for class 0 and 0.67 for class1. The macro and weighted averages of 0.72 for precision, recall, and F1-score further confirm the model's overall effectiveness while suggesting that enhancing the detection of positive cases could boost performance (Figure **1O** and **1P**).

The AdaBoost model has an accuracy of 61.11% and precision is similar for both classes, at 61% for class 0 and 62% for class 1, reflecting balanced prediction rates. However, recall is notably higher for class 0 at 85%, showing that the model effectively identifies non-gestational diabetes cases, but only 31% for class 1, highlighting its struggle to detect gestational diabetes. The F1-score is 0.71 for class 0, demonstrating a good balance between precision and recall, while it drops to 0.42 for class 1, underscoring the model's difficulty with positive cases.

The Naive Bayes model demonstrates accuracy of 72.22% and the precision for GDM is notably high at 88%. However, the recall for GDM is lower at 44%, meaning the model misses a significant portion of actual positive cases. In contrast, the model excels at identifying non GDM with a recall of 95%, but its precision for this class is lower at 68%. The F1-scores reflect this performance: 0.79 for class 0 and 0.58 for class 1 (Figure **1Q** and **1R**). Taken together aforementioned machine learning models showed a diverse range of performance indices.

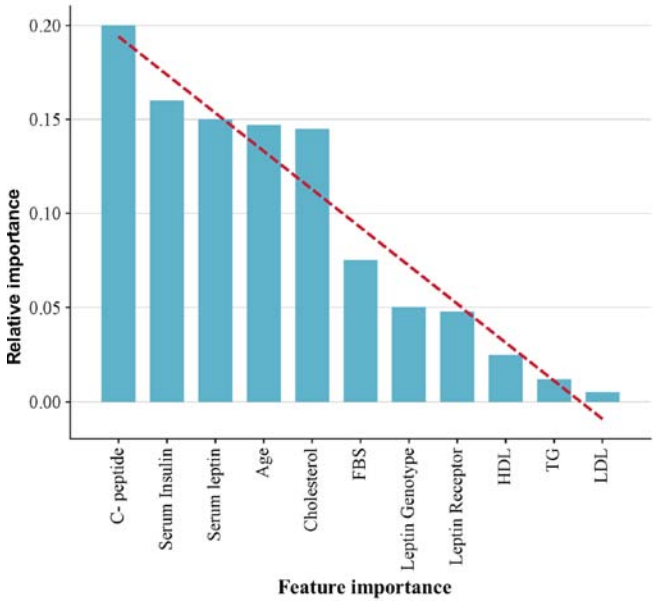

Figure 2: The graph represents feature importance vs relative importance of various indices used to assess GDM risk.

3.2. C-peptide Emerged as the Most Significant Feature for Predicting Gestational Diabetes Mellitus (GDM) using Machine Learning Models

The relative importance of the features used for the machine learning algorithms shows that C-peptide is the most important, followed by fasting insulin, serum leptin levels, age, cholesterol, fasting blood glucose, leptin and LepR genotypes, HDL, TG and LDL in the decreasing order of importance in predicting GDM (Figure **2**).

3.3. Different Machine Learning Models Yield Varying AUC (Area under the Curve) Values, Highlighting their Distinct Strengths and Differences in Predicting GDM

Logistic Regression (AUC = 0.821): This model performs well, with an AUC indicating good predictive accuracy. It suggests that logistic regression is effective in distinguishing between classes based on the given features. Decision Tree (AUC = 0.72): This model has a lower AUC, suggesting that decision trees may not capture complex relationships in the data as effectively as some other algorithms. Gradient Boosting (AUC = 0.76): Gradient boosting performs reasonably well and is better than decision trees, showing it can effectively improve model performance through ensemble methods. Support Vector Machine (SVM) (AUC = 0.74): SVM performs better than decision trees and is on par with Naive Bayes and Neural Networks. It indicates SVM is fairly good at classification but not the best among the models tested-Nearest Neighbors (KNN) (AUC = 0.64): KNN has the lowest AUC, indicating it's the least effective model among those tested for this particular problem. KNN may struggle with high-dimensional data or noisy features. Neural Network (AUC = 0.70): Neural networks show moderate performance. They perform better than KNN but not as well as some of the other models like logistic regression and random forest. AdaBoost (AUC = 0.68): AdaBoost performs worse compared to many other models. While it can be effective in boosting the performance of weak learners, its effectiveness here is limited. Naive Bayes (AUC = 0.74): Naive Bayes performs similarly to SVM, suggesting it's reasonably good but not the top performer. Random Forest (AUC = 0.80): Random Forest shows strong performance, similar to logistic regression, and is better than most other models except for XGBoost. XGBoost (AUC = 0.81): XGBoost achieves the highest AUC among all models tested, indicating it is the most effective at distinguishing between classes based on the provided features. It handles complex data and interactions well (Figure **3A** and **3B**).

3.4. Leptin Gene Polymorphism's Predictive Ability for GDM is Evidenced by the Diverse AUC Values Obtained from Various Machine Learning Models

The results of various predictive models for using Leptin gene polymorphism to predict gestational diabetes show a range of performance metrics. Among these, Logistic Regression emerged as the most effective model with an accuracy of 75.93%, a precision of 76.47%, a recall of 59.09%, and an F1 score of 66.67%. This indicates a good balance between precision and recall, suggesting the model is reliable in predicting true positive cases of gestational diabetes. The Decision Tree model follows closely with an accuracy of 74.07% and a precision of 72.22%, but it shares the same recall as Logistic Regression,

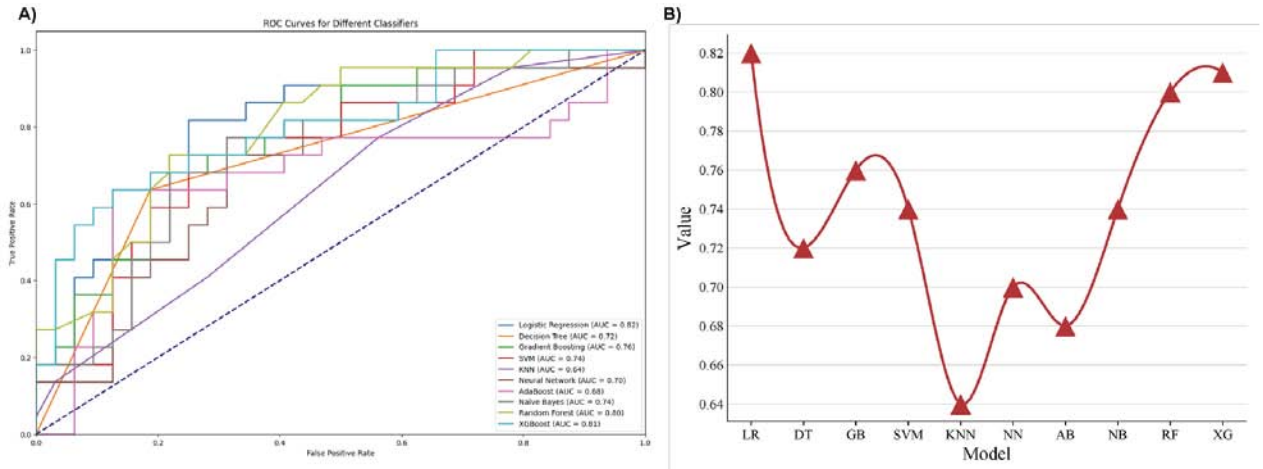

Figure 3: A) The graph represents true positive rate vs false positive rate of different machine learning algorithms. Each colour indicates different machine learning model. **B**) The graph represents area under curve of various ML models obtained after ROC analysis. x – axis represents ML models and y- axis represents AUC.

resulting in a slightly lower F1 score of 65.00%. The Random Forest and Gradient Boosting models both show a decrease in performance, particularly in recall, with accuracies of 68.52% and F1 scores of 48.48% and 54.05%, respectively. AdaBoost and SVM models each have an accuracy of 66.67% and exhibit similar precision and recall values, leading to identical F1 scores of 47.06%. The K-Nearest Neighbors and Neural Network models demonstrate the lowest accuracies of 59.26% and exhibit lower precision and recall values, resulting in an F1 score of 45.00%. Naive Bayes, while having the same accuracy as AdaBoost and SVM, shows slightly better precision but similar recall, leading to a moderate F1 score of 50.00%. Overall, Logistic Regression and Decision Tree models appear to be the most promising for predicting gestational diabetes using Leptin gene polymorphism, whereas the other models show varying degrees of lower effectiveness (Figure **4**).

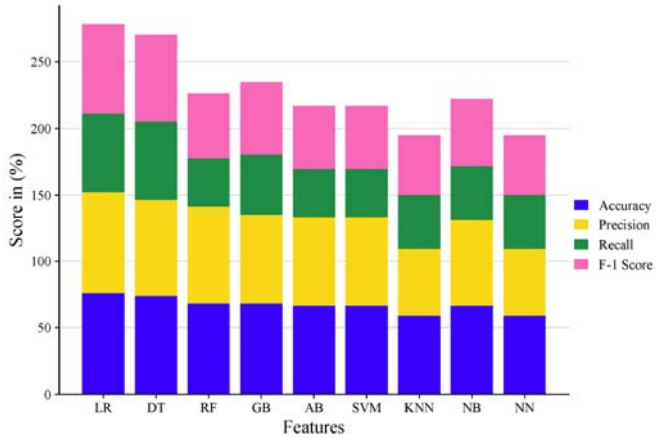

Figure 4: The graph represents Accuracy, precision, Recall and F-1 scores of various ML models considering LEPTIN gene polymorphisms. The x - axis shows ML models and y – axis represents percentage.

3.5. The Predictive Potential of Leptin Receptor Gene Polymorphisms for GDM is Underscored by the Varying AUC Values Produced Across Different Machine Learning Models

The results for models predicting GDM using Leptin receptor gene polymorphism indicate varying levels of effectiveness. Logistic Regression and Neural Network models both achieved the highest accuracy of 64.81%, with a precision of 80.00%, recall of 18.18%, and F1 score of 29.63%. These models exhibit high precision but low recall, indicating they are good at identifying true positives but miss a significant number of actual cases. Decision Tree, Random Forest, and Gradient Boosting models all have an accuracy of 61.11%, precision of 57.14%, recall of 18.18%, and an F1 score of 27.59%. These models also show low recall, meaning they are less effective at identifying all true positive cases. The AdaBoost, SVM, and Naive Bayes models, with accuracies of 59.26%, failed to make any correct positive predictions, resulting in a precision, recall, and F1 score of 0.00%. This indicates these models are not suitable for predicting GDM based on Leptin receptor gene polymorphism in this dataset. The K-Nearest Neighbors (KNN) model has the lowest accuracy at 57.41%, with a precision of 44.44%, recall of 18.18%, and an F1 score of 25.81%.Overall, while Logistic Regression and Neural Network models perform relatively better, their low recall suggests limitations in identifying all cases of GDM. Further model tuning or the incorporation of additional features might be necessary to improve predictive performance (Figure **5**).

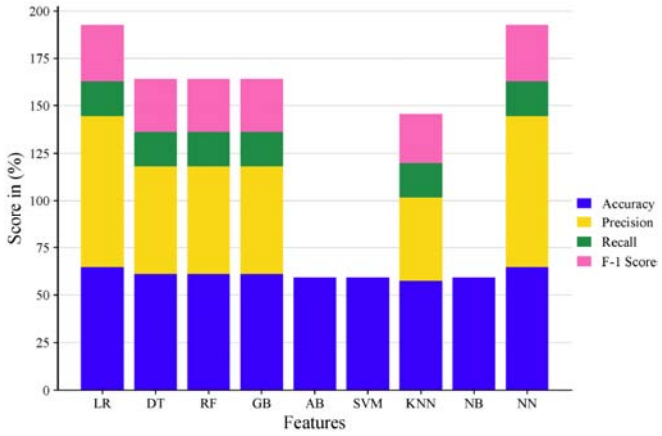

Figure 5: The graph represents Accuracy, precision, Recall and F-1 scores of various ML models considering leptin receptor polymorphisms. The x - axis shows ML models and y – axis represents percentage.

Based on the analysis of the ROC curves, the models displayed a range of performance as indicated by their Area Under Curve (AUC) values. The K-Nearest Neighbors (KNN) model had the lowest AUC at 0.84, followed by the Decision Tree with an AUC of 0.86. The Naïve Bayes model demonstrated a slightly better performance with an AUC of 0.89. Both logistic regression and Support Vector Machine (SVM) models showed similar predictive capabilities with an AUC of 0.91, closely matched by the Gradient Boosting model with an AUC of 0.92. The Random Forest model outperformed these with an AUC of 0.93. Among all the models, the highest predictive accuracy was achieved by the XGBoost model, which recorded the highest AUC of 0.94, indicating its potential as the most effective model for predicting the onset of Gestational Diabetes Mellitus (GDM). Thus, XGBoost emerged as the superior model, while other models such as

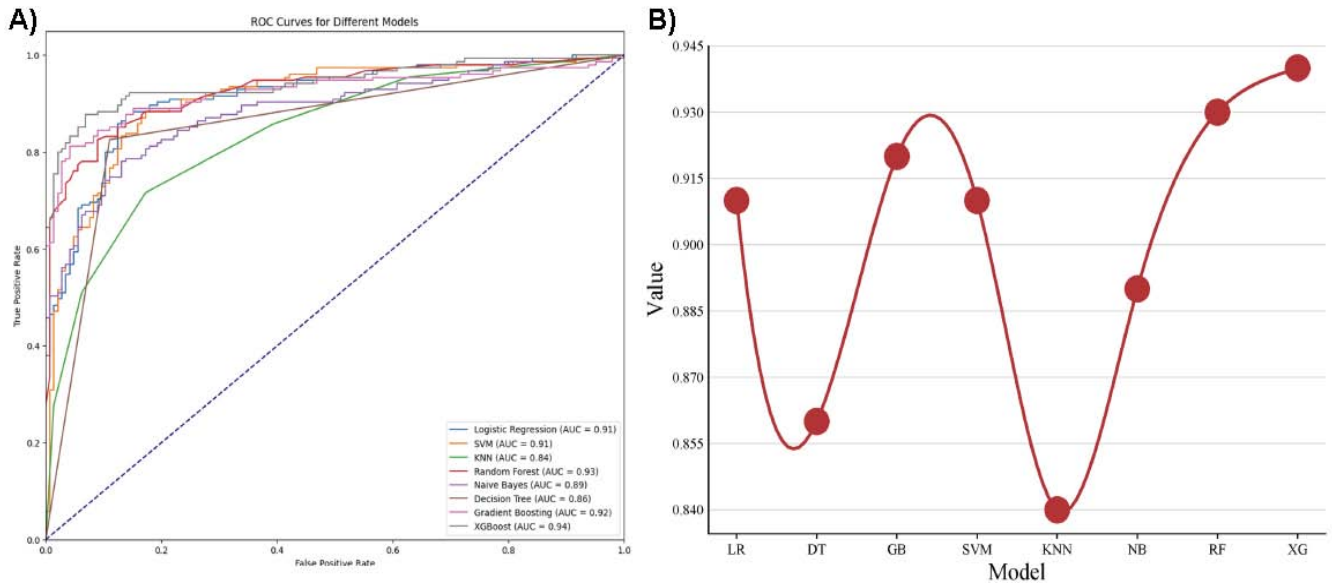

Figure 6: A) The graph represents true positive rate vs false positive rate of different machine learning algorithms. Each colour indicates different machine learning model. **B**) The graph represents area under curve of various ML models obtained after ROC analysis. x – axis represents ML models and y- axis represents AUC.

Random Forest and Gradient Boosting fell between the top performer and the lower-performing KNN and Decision Tree models.

4. DISCUSSION

The study used biomarkers like leptin and LepR genotypes and other clinical variables to construct strong machine learning models to predict gestational diabetes mellitus (GDM). Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, SVM, KNN, Neural Network, XGBoost, AdaBoost, and Naive Bayes were examined. These models were assessed for accuracy, precision, recall, and F1-score.The best model was Logistic Regression, with 75% accuracy. It reliably predicted genuine positive GDM patients with excellent accuracy (76%), recall (75%), and F1-score (74%). However, the model had a propensity to produce more false negatives for GDM instances, resulting in worse class 1 recall. With 69% accuracy, the Decision Tree model worked well. It exhibited balanced precision and recall for non-GDM instances but a greater rate of false negatives for GDM cases, resulting in a modest F1-score for class 1. This model may benefit from tweaking or ensemble approaches to improve prediction. The robust Random Forest model predicted GDM with 61% accuracy, showing a large imbalance. The model has excellent accuracy but low recall for GDM cases, resulting in many missed cases. Class imbalance must be addressed to improve GDM detection. Gradient Boosting performed moderately with 67% accuracy. It exhibited balanced accuracy for both classes, but low recall for GDM, resulting in a lower F1-score for class 1. The model may be better at recognizing non-GDM situations than GDM cases. With 56% accuracy, the SVM model performed badly. Class 1 recall was 0%, indicating no GDM predictions. This model cannot recognize GDM situations, making it unsuitable for this scenario. Overall accuracy for KNN was 58%, with acceptable precision for GDM but low recall, resulting in many false negatives. To accurately anticipate GDM, this model needs major upgrades. The Neural Network model has 58% accuracy and balanced precision and recall for non-GDM instances. GDM identification was difficult, as seen by its lower class 1 F1-score. XGBoost performed well with 72% accuracy. It had balanced accuracy and recall for both classes and an excellent GDM F1-score. One of the most effective methods in this study is this model. AdaBoost exhibited 61% accuracy, balanced precision for both classes, and minimal GDM recall. Its inability to recognize affirmative cases suggests additional refining. Naive Bayes had 72% accuracy, good GDM precision, and low recall. This approach predicted true positives effectively but missed several GDM cases. Model optimization and class imbalances are needed due to model performance differences. The most accurate models are Logistic Regression and XGBoost, which balance accuracy and recall.

Several independent studies have shown that, machine learning models were reliable in predicting the risk of GDM. For example, a study by Rustam *et al*., showed that machine learning techniques surpassed

conventional statistical methods in forecasting gestational diabetes mellitus (GDM), with the CatBoost Classifier attaining the greatest accuracy of 93% by utilizing improved predictor factors. This study emphasizes the capability of sophisticated algorithms for enhanced GDM risk assessment [8]. An review by Lu *et al.*, also demonstrated the potential role of machine learning algorithms for the better management of the GDM [9]. A clinical diagnostic system utilizing deep learning through recurrent neural network — long short-term memory (RNN-LSTM) and Bayesian optimization attained 95% sensitivity, 99% specificity, and 98% AUC in detecting gestational diabetes (GD) risk, therefore minimizing needless oral glucose tolerance testing and conserving time and resources [10]. A Japanese study has revealed that features such as anthropometry, maternal birthweight, early pregnancy, lifestyle and socioeconomic status in relation to pregnancy (1^{st} trimester) has shown that, gradient boosting decision tree (GBDT) has shown a strong association with GDM. However, the authors reported that The AUC for individuals with a history of GDM was poor (0.67); however, including maternal genetic data in future models is anticipated to improve predictive accuracy [11]. The authors [11] noted that an AUC of 0.70 or above is deemed acceptable for clinical implications, and most of our machine learning models predicted using Leptin and Leptin R exhibited values over 0.70, signifying the clinical significance of our findings. Notably, the Support Vector Machine (SVM) classifier showed particularly low recall for GDM cases. This outcome may be attributed to several factors: (1) the relatively small sample size and class imbalance, which can bias SVM decision boundaries toward the majority class; (2) the heterogeneous and potentially non-linear relationships among genetic and biochemical predictors, which may not be fully captured without extensive kernel tuning; and (3) limited hyperparameter optimization in the current analysis. These considerations highlight the importance of data balancing strategies and parameter optimization when applying SVM to clinical datasets of this nature.

Recent research highlights the drawbacks of OGTT, including as logistical issues and patient discomfort [12-14]. For instance, studies have emphasized how new biomarkers, such as leptin, provide early prediction capabilities with less invasiveness [15, 16]. Further, combining genetic information with clinical characteristics has potential for lowering OGTT dependence, especially in high-risk populations [17]. An independent investigation revealed that elevated circulation levels of leptin (OR: 1.16, 95% CI: 1.07–1.27), IL-6 (OR: 1.35, 95% CI: 1.05–1.73), and TNF-α (OR: 1.28, 95% CI: 1.01–1.62) were substantially correlated with an augmented risk of gestational diabetes mellitus (GDM). The findings indicated that the examined cytokines may function as possible biomarkers for the etiology of GDM, necessitating further extensive longitudinal research for confirmation [18]. A randomized control trail by Ramos *et al.*, demonstrated that, 40 SNPs substantially correlated with gestational diabetes mellitus (GDM) in Caucasian and Hispanic women, emphasizing genetic variations associated with elevated or reduced GDM risk. The results indicate that the incorporation of genetic markers in the prediction of the GDM along with the traditional markers [19]. One more variable of interest in predicting the GDM is C-peptide. A separate research indicated that women with a history of prenatal diabetes mellitus (pGDM) who exhibited fasting plasma glucose levels over 5 mmol/L during pregnancy and 12 weeks postpartum had a markedly elevated risk of acquiring type 2 diabetes within 8 to 10 years. Furthermore, elevated C-peptide levels and decreased ghrelin levels postpartum were identified as major risk factors [20].

Overall, the current study's findings align with existing literature, affirming the reliability of models like Logistic Regression and XGBoost for GDM prediction, and the significance of biomarkers such as HOMA-IR and C peptide levels. However, the consistent challenges faced by models like Decision Tree, Random Forest, SVM, and KNN in handling class imbalances and false negatives indicate a need for continued refinement and tuning. These insights are crucial for enhancing the effectiveness of machine learning models in clinical settings for the prediction and management of GDM.

The novel screening tool effectively identifies individuals at risk for GDM, with performance metrics comparable to machine learning models and previous studies. Machine learning models offer higher sensitivity, making them valuable for early detection; however, their lower specificity may lead to over-screening. The significant association of HOMA-IR and C peptide levels with GDM suggests their importance as predictive biomarkers. Future studies should consider these markers to enhance predictive accuracy. Additionally, incorporating ensemble

methods and improving model interpretability could further aid healthcare professionals in early detection and management of GDM.

5. LIMITATIONS OF THE STUDY

The cost of genotyping has constantly decreased making it more practical for large-scale clinical applications [21, 22]. Furthermore, recent research shows that integrating genetic and clinical data considerably improves early GDM prediction, outweighing logistical constraints [23-26]. This research represents a substantial advancement by including genetic variations for predictive analysis. However, owing to its constrained sample size, emphasis on a particular trimester, more validation is required prior to generalizing these findings to wider clinical contexts. Thus, while genetic data usage is not yet standard, our study provides a forward-looking framework for integrating these technologies. We used a single train–test split without formal cross-validation or specific strategies to address potential class imbalance. While this straightforward approach allowed for initial model evaluation, it may limit the robustness and generalizability of our findings. Future studies should incorporate cross-validation and dedicated imbalance-mitigation methods to enhance model validity and reliability.

6. FUTURE DIRECTIONS

Optimizing models through fine-tuning hyperparameters and exploring ensemble techniques can significantly enhance performance, particularly for gestational diabetes mellitus (GDM) detection. Addressing class imbalance by applying methods like SMOTE (Synthetic Minority Over-sampling Technique) can improve the recall of GDM cases, ensuring a more balanced dataset. Additionally, incorporating a broader range of potential biomarkers may increase the models' predictive power. Finally, developing interpretable models is crucial, as it ensures that healthcare professionals can effectively utilize these tools for early prediction and intervention, ultimately improving patient outcomes.

7. CONCLUSION

This study leverages the power of machine learning to predict gestational diabetes mellitus using clinical features and biomarkers, particularly focusing on the genotype of leptin and leptin receptor. By comparing various algorithms, the study seeks to identify the most effective and interpretable model that can aid in the early detection and management of GDM. This approach aims to improve maternal and foetal health outcomes by enabling timely interventions and personalized treatment plans.

FUNDING SOURCES

Nil.

ACKNOWLEDGEMENTS

The authors thank Central Research Laboratory for Molecular Genetics, Bioinformatics and Machine Learning at Apollo Institute of Medical Sciences and Research Chittoor, Andhra Pradesh, India.

CONFLICTS OF INTEREST OF EACH AUTHOR

The authors declare no potential conflicts of interest.

AUTHORSHIP STATEMENT

The manuscript has been read and approved by all the authors.

LIST OF ABBREVIATIONS

GDM	=	Gestational Diabetes Mellitus
OGTT	=	Oral Glucose Tolerance Test
PPV	=	Positive Predictive Value
NPV	=	Negative Predictive Value
BMI	=	Body Mass Index
LepR	=	Leptin Receptor
ADA	=	American Diabetes Association
PIH	=	Pregnancy-Induced Hypertension
DNA	=	Deoxyribonucleic Acid
EDTA	=	Ethylenediaminetetraacetic Acid
PCR	=	Polymerase Chain Reaction
RFLP	=	Restriction Fragment Length Polymorphism
TG	=	Triglycerides

LDL = Low-Density Lipoprotein

HDL = High-Density Lipoprotein

SVM = Support Vector Machine

KNN = K-Nearest Neighbors

ROC = Receiver Operating Characteristic

AUC = Area Under Curve

HOMA-IR = Homeostatic Model Assessment for Insulin Resistance

REFERENCES

[1] American Diabetes Association. 14. Management of diabetes in pregnancy: standards of medical care in diabetes—2021. Diabetes Care 2020; 44(Suppl 1): S200-S210. https://doi.org/10.2337/dc21-S014

[2] Wang H, *et al.* IDF Diabetes Atlas: estimation of global and regional gestational diabetes mellitus prevalence for 2021 by International Association of Diabetes in Pregnancy Study Group's criteria. Diabetes Res Clin Pract 2022; 183: 109050. https://doi.org/10.1016/j.diabres.2021.109050

[3] Yoon KH, *et al.* Epidemic obesity and type 2 diabetes in Asia. Lancet 2006; 368(9548): 1681-1688. https://doi.org/10.1016/S0140-6736(06)69703-1

[4] Nguyen CL, Pham NM, Binns CW, Van Duong D, Lee AH. Prevalence of gestational diabetes mellitus in Eastern and Southeastern Asia: a systematic review and meta-analysis. J Diabetes Res 2018; 2018: 6536974. https://doi.org/10.1155/2018/6536974

[5] Bimson BE, Rosenn BM, Morris SA, Sasso EB, Schwartz RA, Brustman LE. Current trends in the diagnosis and management of gestational diabetes mellitus in the United States. J Matern Neonatal Med 2017; 30(21): 2607-2612. https://doi.org/10.1080/14767058.2016.1257603

[6] Gallardo-Rincón H, *et al.* Diagnostic accuracy of capillary blood glucometer testing for gestational diabetes. Diabetes Metab Syndr Obes 2022; 15: 3855-3870. https://doi.org/10.2147/DMSO.S389420

[7] Adiga U, Banawalikar N, Rai T. Association of leptin and leptin receptor gene polymorphisms with insulin resistance in pregnant women: a cross-sectional study. F1000Research 2022; 11: 692. https://doi.org/10.12688/f1000research.122537.2

[8] Rustam F, *et al.* Enhanced detection of diabetes mellitus using novel ensemble feature engineering approach and machine learning model. Sci Rep 2024; 14(1): 23274. https://doi.org/10.1038/s41598-024-74357-w

[9] Lu HY, *et al.* Digital health and machine learning technologies for blood glucose monitoring and management of gestational diabetes. IEEE Rev Biomed Eng 2024; 17: 98-117. https://doi.org/10.1109/RBME.2023.3242261

[10] Kurt B, *et al.* Prediction of gestational diabetes using deep learning and Bayesian optimization and traditional machine learning techniques. Med Biol Eng Comput 2023; 61(7): 1649-1660. https://doi.org/10.1007/s11517-023-02800-7

[11] Watanabe M, *et al.* Prediction of gestational diabetes mellitus using machine learning from birth cohort data of the Japan Environment and Children's Study. Sci Rep 2023; 13(1): 17419. https://doi.org/10.1038/s41598-023-44313-1

[12] Lachmann EH, Fox RA, Dennison RA, Usher-Smith JA, Meek CL, Aiken CE. Barriers to completing oral glucose tolerance testing in women at risk of gestational diabetes. Diabet Med 2020; 37(9): 1482-1489. https://doi.org/10.1111/dme.14292

[13] Bogdanet D, O'Shea P, Lyons C, Shafat A, Dunne F. The oral glucose tolerance test—Is it time for a change?—A literature review with an emphasis on pregnancy. J Clin Med 2020; 9(11): 3451. https://doi.org/10.3390/jcm9113451

[14] Kuo FY, Cheng K-C, Li Y, Cheng J-T. Oral glucose tolerance test in diabetes, the old method revisited. World J Diabetes 2021; 12(6): 786-793. https://doi.org/10.4239/wjd.v12.i6.786

[15] Inthavong S, Jatavan P, Tongsong T. Predictive utility of biochemical markers for the diagnosis and prognosis of gestational diabetes mellitus. Int J Mol Sci 2024; 25(21): 11666. https://doi.org/10.3390/ijms252111666

[16] Omazić J, *et al.* Early markers of gestational diabetes mellitus: what we know and which way forward? Biochem Med (Zagreb) 2021; 31(3): 30502. https://doi.org/10.11613/BM.2021.030502

[17] Ruchat S-M, Vohl M-C, Weisnagel SJ, Rankinen T, Bouchard C, Pérusse L. Combining genetic markers and clinical risk factors improves the risk assessment of impaired glucose metabolism. Ann Med 2010; 42(3): 196-206. https://doi.org/10.3109/07853890903559716

[18] Hosseini E, Mokhtari Z, Salehi Abargouei A, Mishra GD, Amani R. Maternal circulating leptin, tumor necrosis factor-alpha, and interleukine-6 in association with gestational diabetes mellitus: a systematic review and meta-analysis. Gynecol Endocrinol 2023; 39(1): 2183049. https://doi.org/10.1080/09513590.2023.2183049

[19] Ramos-Levi A, *et al.* Genetic variants for prediction of gestational diabetes mellitus and modulation of susceptibility by a nutritional intervention based on a Mediterranean diet. Front Endocrinol (Lausanne) 2022; 13: 1036088. https://doi.org/10.3389/fendo.2022.1036088

[20] Lappas M, Jinks D, Ugoni A, Louizos CCJ, Permezel M, Georgiou HM. Post-partum plasma C-peptide and ghrelin concentrations are predictive of type 2 diabetes in women with previous gestational diabetes mellitus. J Diabetes 2015; 7(4): 506-511. https://doi.org/10.1111/1753-0407.12209

[21] Bianconi I, Aschbacher R, Pagani E. Current uses and future perspectives of genomic technologies in clinical microbiology. Antibiotics (Basel) 2023; 12(11): 1580. https://doi.org/10.3390/antibiotics12111580

[22] Nafea AM, *et al.* Application of next-generation sequencing to identify different pathogens. Front Microbiol 2023; 14: 1329330. https://doi.org/10.3389/fmicb.2023.1329330

[23] Tian Y, Li P. Genetic risk score to improve prediction and treatment in gestational diabetes mellitus. Front Endocrinol (Lausanne) 2022; 13: 955821. https://doi.org/10.3389/fendo.2022.955821

[24] Zhang C, *et al.* Genetic variants and the risk of gestational diabetes mellitus: a systematic review. Hum Reprod Update 2013; 19(4): 376-390. https://doi.org/10.1093/humupd/dmt013

[25] Wu Q, *et al.* An early prediction model for gestational diabetes mellitus based on genetic variants and clinical characteristics in China. Diabetol Metab Syndr 2022; 14(1): 15. https://doi.org/10.1186/s13098-022-00788-y

[26] Kaya Y, Bütün Z, Çelik Ö, Salik EA, Tahta T, Yavuz AA. The early prediction of gestational diabetes mellitus by machine learning models. BMC Pregnancy Childbirth 2024; 24(1): 574. https://doi.org/10.1186/s12884-024-06783-7

Received on 09-08-2025 Accepted on 03-09-2025 Published on 26-09-2025

https://doi.org/10.6000/1929-6029.2025.14.55