UtilitasMathematica

ISSN 0315-3681 Volume 122 (2), 2025

Overview of Machine and Deep Learning Models for Gestational Diabetes Mellitus (GDM) Prediction

[1] NUTHALAPATI HIMAJA, [2] Dr.HARI KIRAN VEGA, [3] Dr.D.NAGA MALLESWARI

Department of CSE

Koneru Lakshmaiah Education Foundation, Green Fields Vaddeswaram, Andhra Pradesh, India

E-MAIL: [1]himajanuthalapati2001@gmail.com , [2]hari.vega@kluniversity.in, [3]nagamalleswary@kluniversity.in

***Abstract*— Gestational Diabetes Mellitus (GDM) is a form of diabetes that develops during pregnancy, posing significant risks to both mother and child. Although conventional diagnostic tests are reliable, they are often time-consuming and invasive. To address these limitations, recent studies have increasingly turned to data-driven predictive methods. Machine Learning (ML) algorithms, including Decision Trees, Random Forests, and XGBoost, have demonstrated strong potential in early GDM identification. Furthermore, advanced Deep Learning (DL) models like Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) have shown superior capability in modelling complex temporal health data. Hybrid architectures, such as GRU-LSTM, have reported predictive accuracies as high as 99%. Despite these promising outcomes, challenges persist in addressing class imbalance and effective data preprocessing. Comparative evaluations highlight trade-offs between predictive performance, interpretability, and computational efficiency. Future research should focus on leveraging diverse datasets, integrating explainable AI (XAI) methods, and enhancing clinical integration to support the real-world deployment of robust GDM prediction systems.**

Key words: Machine Learning, GRU, Deep Learning, LSTM, SMOTE, Prediction Models, Gestational Diabetes Mellitus.

I. Introduction

Diabetes Mellitus (DM) is widely recognised as one of the most pressing global health challenges, affecting hundreds of millions of individuals across all age groups and regions. The disease not only imposes a heavy burden on patients and healthcare systems but also contributes to life-threatening complications, including cardiovascular disease, kidney failure, neuropathy, and vision impairment such as diabetic retinopathy. Amongst the different types of diabetes, Gestational Diabetes Mellitus (GDM) has gained special attention due to its occurrence during pregnancy in women without any prior history of diabetes. GDM is considered a unique subtype of DM, yet its consequences extend well beyond the pregnancy period, impacting both maternal and neonatal health.

For expectant mothers, GDM can lead to severe complications such as preeclampsia, gestational hypertension, and an increased probability of caesarean section deliveries. On the infant side, risks include neonatal hypoglycaemia, macrosomia (abnormally high birth weight), respiratory distress, and a heightened susceptibility to obesity and Type 2 diabetes later in life. Thus, timely prediction and early detection of GDM are crucial for ensuring maternal well-being and minimising short- and long-term health risks for newborns.

Traditionally, diagnosis of GDM relies on fasting blood glucose (FBG) measurements and the oral glucose tolerance test (OGTT). While clinically reliable, these tests come with notable drawbacks: they are invasive, costly, and time-consuming, making them unsuitable for frequent screening or large-scale population monitoring. Furthermore, they only detect GDM once abnormal glucose regulation has already occurred, rather than offering predictive insights before the onset of complications. These limitations highlight the pressing need for innovative, non-invasive, and scalable approaches to prediction and early diagnosis.

In recent years, machine learning (ML) and deep learning (DL) have become powerful enablers in healthcare data analytics [1, 2, 4, 16]. By leveraging large-scale clinical, demographic, and lifestyle information, these approaches are able to uncover complex and non-linear relationships that often remain hidden in conventional statistical methods. For the prediction of gestational diabetes mellitus (GDM), researchers have experimented with a diverse set of ML techniques, including Logistic Regression, Decision Trees, Random Forests, Support Vector Machines (SVM), K-Nearest Neighbours (KNN), and advanced ensemble methods such as XGBoost [3, 4, 6, 8, 11].

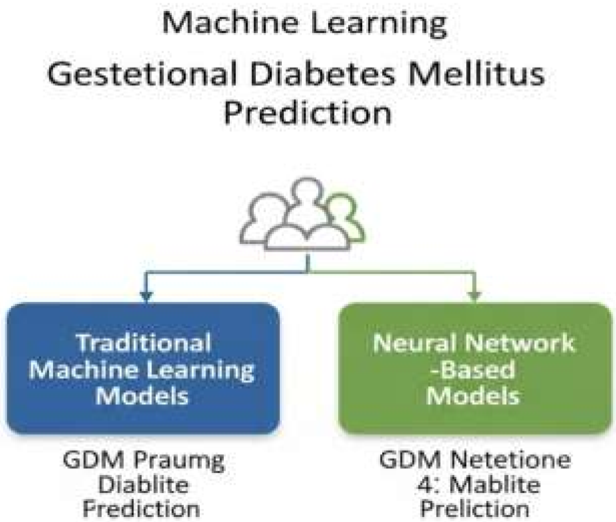
At the same time, deep learning architectures—particularly Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) models, and Gated Recurrent Units (GRUs)—have shown strong capability in handling sequential or time-series data, such as glucose monitoring trends, insulin profiles, and blood pressure changes throughout pregnancy [7, 9].

More recently, hybrid deep learning models that integrate both GRU and LSTM elements have achieved highly competitive performance, with reported accuracies nearing 99% [9]. These models capitalise on the long-term memory strength of LSTM while benefiting from the computational efficiency of GRU, thus providing robust detection of intricate temporal dependencies in patient datasets. However, despite these advancements, several challenges remain unresolved. Common issues include class imbalance within datasets, poor generalisability across diverse patient populations, and the considerable computational demand of training such models [2, 3]. To address these concerns, strategies such as the Synthetic Minority Oversampling Technique (SMOTE) and advanced feature engineering have been applied, though additional refinement is still necessary [3].

Within this context, the purpose of this review is to examine the current landscape of ML and DL approaches for GDM prediction, assess their respective advantages and limitations, identify existing research gaps, and outline potential avenues for improvement. The overarching aim is to contribute towards the development of reliable, scalable, and efficient predictive systems that can support healthcare professionals in the early identification of GDM, ultimately enhancing both maternal and neonatal health outcomes.

II. THE ROLE OF MACHINE LEARNING IN DIABETES PREDICTION

In predictive healthcare, machine learning (ML) has emerged as a transformative technology. It can analyse large datasets to uncover intricate patterns that traditional statistical methods might miss [1, 4]. Within the framework of Gestational Diabetes Mellitus (GDM) prediction, ML offers a data-driven model that can be used to select high-risk patients in the initial stage, hence, allowing them to take timely measures to prevent maternal and fatal issues [3, 11]. ML approaches for GDM prediction are generally categorised into traditional machine learning classifiers and neural network-based architectures, each offering unique advantages and limitations depending on the dataset characteristics and clinical requirements.



A. Traditional Machine Learning Models

Traditional machine learning algorithms such as Decision Trees (DT), Random Forests (RF), Extreme Gradient Boosting (XGBoost), K-Nearest Neighbours (KNN), Support Vector Machines (SVM), and Logistic Regression are frequently utilised in the prediction of gestational diabetes mellitus (GDM)[1, 3, 6, 8, 11] .These methods remain popular because they offer a balance of interpretability, robustness, and relatively low computational cost compared to more complex deep learning approaches.

Decision Trees (DT) are simple and easy-to-understand models that partition data based on certain feature thresholds [6]. A major drawback, however, is their tendency to overfit, particularly with smaller datasets [8]. Random Forest, which combines multiple decision trees into an ensemble, mitigates overfitting by averaging predictions from individual trees, enhancing stability and generalisation [3, 8, 13]. Logistic Regression serves as a fundamental model, appreciated for its simplicity and ease of interpretation, although it may struggle with non-linear patterns inherent in medical data.

B. Neural Network-Based Models

Neural networks offer a robust alternative for modelling complex, non-linear relationships [9]. Recurrent Neural Networks (RNNs) are particularly effective at processing sequential clinical data but can encounter issues such as vanishing gradients [7, 9]. Long Short-Term Memory (LSTM) networks overcome these limitations by retaining long-term dependencies, making them well-suited for sequential and temporal health datasets [7, 9].

Table 1: Comparative Overview of Machine Learning Models for GDM Prediction

Model	Accuracy Range	Advantages	Limitations
Decision Tree (DT)	70–85%	Highly interpretable, simple to implement	Prone to overfitting, sensitive to small data variations
Random Forest (RF)	85–97%	Reduces overfitting, robust, good generalization	Less interpretable than single DT, computationally heavier
XGBoost	Up to 91%	High accuracy, handles missing data well	Complex to tune, computationally intensive
K-Nearest Neighbors (KNN)	70–82%	Simple, effective on balanced datasets	Poor performance in high-dimensional data, sensitive to noise
Support Vector Machine (SVM)	74–89%	Effective on small/medium datasets, good generalization	Poor scalability with large datasets, less interpretable
Logistic Regression	68–80%	Simple, interpretable, fast	Limited for non-linear relationships
Recurrent Neural Network (RNN)	Up to 96%	Captures sequential patterns	Vanishing gradient problem, slower training
(LSTM)	~82%	Handles long term dependencies, suitable for sequences	Computationally expensive, longer training time
Gated Recurrent Unit (GRU)	~90%	Fewer parameters, faster convergence	Slightly less expressive than LSTM

III. Addressing Dataset Challenges

Predicting Gestational Diabetes Mellitus (GDM) using machine learning often faces several challenges related to the characteristics and quality of available datasets. One of the most common issues in GDM prediction is class imbalance [3]. This problem can severely impact a model's reliability and overall performance.

A. The Problem of Class Imbalance

In most GDM datasets, the number of positive instances—representing women diagnosed with gestational diabetes mellitus—is substantially lower than the number of negative instances corresponding to healthy pregnancies [3]. This class imbalance creates a significant challenge for predictive modelling, as traditional machine learning algorithms often exhibit a bias towards the majority class [3]. Consequently, models may achieve high overall accuracy yet perform poorly in correctly identifying the minority class, leading to elevated false-negative rates [3]. Such errors are critical in a clinical setting, since undetected GDM cases may remain untreated, increasing risks for both the mother and the newborn.

To address this issue, researchers commonly employ data balancing strategies, with oversampling being one of the most widely adopted [3]. Amongst these, Synthetic Minority Oversampling Technique (SMOTE) has gained particular prominence [3]. SMOTE generates artificial samples of the minority class by interpolating between existing data points, thereby enriching the dataset with more balanced class representation [3]. Empirical research indicates that the integration of SMOTE into model training significantly enhances the detection of positive GDM cases and improves evaluation metrics such as recall and F1-score [3]. By boosting sensitivity while maintaining specificity, this strategy strengthens the overall clinical reliability of predictive models.

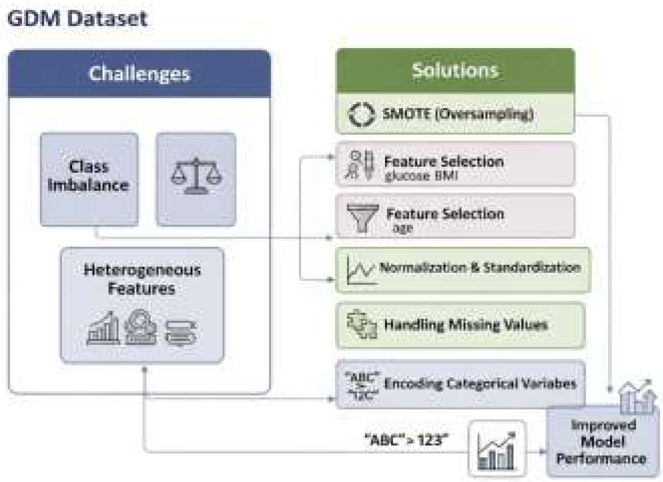
B. Data Preprocessing and Feature Engineering

Preprocessing and feature engineering are essential steps for improving the accuracy and robustness of predictive models [2, 4, 5, 6]. Feature selection is particularly important, as including irrelevant or redundant variables can degrade model performance and increase computational costs [4, 6]. For GDM prediction, selecting the most informative features ensures that models focus on variables with the highest predictive value. Common predictors include glucose levels, body mass index (BMI), maternal age, insulin levels, blood pressure, and pregnancy history [2, 4, 5, 6]. Thoughtful selection of these features enhances both model efficiency and clinical applicability [6].

Table 2: Preprocessing and Feature Engineering Techniques for GDM Prediction

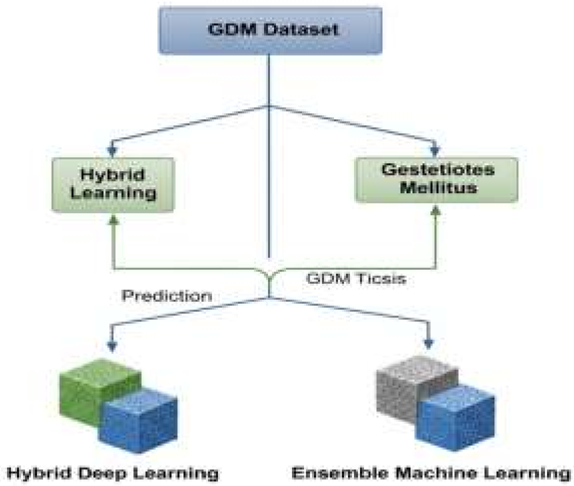
Technique	Purpose	Key Benefits
SMOTE (Oversampling)	Address class imbalance	Improves recall and F1-score for minority class, reduces false negatives
Feature Selection	Identify relevant predictors (glucose, BMI, age, insulin, blood pressure, pregnancy history)	Reduces dimensionality, improves accuracy, decreases computational cost
Normalization	Scale features to a specific range (0–1)	Ensures balanced input for models, speeds up convergence
Standardization	Transform features to zero mean and unit variance	Maintains numerical stability, improves gradient descent performance
Handling Missing Values	Fill or remove missing entries	Prevents bias and improves model reliability
Encoding Categorical Variables	Convert categorical data to numeric form	Enables model compatibility with categorical features

In addition, proper scaling of data is critical, especially for gradient descent-based neural networks like LSTM and GRU [9]. Features with larger numerical ranges can dominate the learning process if not scaled correctly, leading to slower convergence, biased weight assignments, and reduced predictive accuracy. Standardisation and normalisation are widely used techniques to address this [6]. Standardisation adjusts features to have a mean of zero and a unit variance, ensuring all variables contribute equally during training. Normalisation rescales feature values to a defined range, commonly between 0 and 1, promoting numerical stability and faster learning. Together, addressing class imbalance and implementing careful data preprocessing allows machine learning models to predict GDM more reliably, supporting timely clinical decision-making and intervention [3, 6].



IV. THE RISE OF HYBRID AND ENSEMBLE MODELS

With the continued advancement of machine learning (ML) and deep learning (DL), there has been a growing emphasis on hybrid and ensemble approaches to improve the accuracy of Gestational Diabetes Mellitus (GDM) prediction [9, 10, 11]. Hybrid models, like GRU+LSTM networks, merge the distinct benefits of different deep learning architectures. This allows them to identify both short-term and long-term patterns in clinical data at the same time. Ensemble methods, including algorithms like Random Forest and XGBoost, integrate multiple learners to minimise variance and bias, enhancing model robustness and generalisation [1, 11]. By leveraging the strengths of diverse algorithms while offsetting their individual limitations, these strategies provide a more reliable framework for precise and clinically relevant GDM prediction. Such integration not only improves predictive performance but also increases resilience to data variability, making these models particularly suitable for real-world healthcare applications [11, 13].



A. *Hybrid Deep Learning Architectures*

Hybrid deep learning frameworks, especially GRU+LSTM models, have become increasingly popular in analysing sequential medical datasets [9]. By combining the computational efficiency of Gated Recurrent Units (GRU) with the superior long-term memory retention of Long Short-Term Memory (LSTM) networks, these architectures can effectively model both short-term and long-term dependencies in patient information [9]. This dual capability allows the models to detect subtle temporal patterns in clinical parameters—such as glucose measurements, insulin variations, and blood pressure trends—thereby improving early identification of high-risk pregnancies and facilitating timely interventions [9]. The integration of GRU and LSTM also promotes faster convergence during training and better handling of complex sequential data. Research indicates that GRU+LSTM hybrids can achieve predictive accuracies approaching 99%, outperforming single RNN or LSTM models, while offering enhanced stability and generalisation across varied patient populations [9].

B. Ensemble Machine Learning Techniques

Ensemble machine learning techniques have emerged as a powerful strategy to improve the predictive performance of models, particularly in healthcare applications such as Gestational Diabetes Mellitus (GDM) prediction [1, 6]. Instead of relying on a single classifier, ensemble learning combines the strengths of multiple models to produce more reliable, accurate, and generalised outcomes. This approach addresses limitations such as overfitting, variance, and bias, which are often encountered when using individual models on complex and heterogeneous clinical datasets.

There are several widely adopted ensemble methods, including majority voting, bagging, and boosting [13, 14]. In the majority voting method, multiple base learners (such as Decision Trees, Logistic Regression, or Support Vector Machines) generate predictions, and the final classification is determined by the class receiving the most votes. This technique ensures that the collective wisdom of several models reduces the risk of errors introduced by a single weak classifier. For example, when predicting whether a patient is at high risk of developing GDM, using majority voting across multiple base models can provide a more stable and dependable prediction outcome.

Bagging (Bootstrap Aggregating) is another important ensemble technique designed to reduce variance and improve robustness. A well-known example of bagging is the Random Forest algorithm [6, 8], which trains multiple decision trees on randomly sampled subsets of the dataset and then aggregates their outputs. In the context of GDM prediction, Random Forest models have been widely employed due to their ability to handle missing data, nonlinear relationships, and high-dimensional clinical features such as blood glucose levels, BMI, family history, and demographic characteristics [6, 8]. By reducing the risk of overfitting and improving generalisation, bagging-based ensembles provide dependable performance even in noisy datasets [8].

Boosting methods, on the other hand, aim to sequentially correct the weaknesses of previous learners [11]. Algorithms such as XGBoost (Extreme Gradient Boosting), AdaBoost, and LightGBM have shown excellent results in medical classification tasks [6, 11]. XGBoost, in particular, is highly efficient at handling imbalanced datasets—a common issue in GDM studies, where positive cases are relatively fewer compared to non-GDM cases [11]. By assigning higher weights to misclassified samples in each iteration, boosting ensures that the model gradually improves its ability to identify difficult or rare cases. This makes boosting particularly valuable in clinical prediction tasks where missing a high-risk case could have severe implications [6, 11].

Recent studies have also explored the integration of diverse models within ensemble frameworks [13]. For instance, combining XGBoost, K-Nearest Neighbours (KNN), and Random Forest into a hybrid ensemble has demonstrated improved predictive performance compared to using any of these models individually [13]. Such integrated ensembles take advantage of the complementary strengths of different algorithms: while KNN excels at local similarity-based predictions, Random Forest captures non-linear relationships, and XGBoost improves overall sensitivity and accuracy [13]. This diversity enables ensemble approaches to handle the complex, multidimensional nature of healthcare data more effectively.

C. *Comparison and Insights*

Hybrid deep learning models generally demonstrate superior predictive accuracy compared to traditional ensemble machine learning (ML) methods, particularly when working with sequential or time-series data such as patient medical histories [9]. These models, by integrating architectures like GRU and LSTM, are capable of capturing both short-term and long-term temporal dependencies, which are often critical in accurately predicting the onset of Gestational Diabetes Mellitus (GDM) [9]. Their ability to model complex nonlinear relationships enables them to achieve highly reliable results, often surpassing 99% accuracy in experimental studies [9].

Despite their impressive performance, hybrid deep learning approaches tend to require substantial computational power, large datasets for training, and extended training times, which can make them less practical in low-resource healthcare settings [9]. In contrast, ensemble ML methods—such as majority voting, bagging, and boosting techniques like XGBoost combined with KNN and Random Forest—offer a more computationally efficient alternative [6, 8, 11]. While these models may achieve slightly lower accuracy levels (around 86%), they are easier to implement, faster to train, and more adaptable to real-world clinical environments where resources such as high-performance GPUs may not be readily available.

The choice between hybrid deep learning and ensemble ML approaches therefore hinges on multiple factors, including the size and nature of the dataset, the computational resources available, and the urgency of deployment within clinical workflows [6]. For instance, in tertiary care hospitals with advanced infrastructure, hybrid DL models may be more suitable due to their unparalleled accuracy [9]. However, in rural or resource-constrained healthcare environments, ensemble methods provide a more feasible and cost-effective solution while still maintaining clinically acceptable performance [6, 8, 11].

Table 3: Comparative Overview of Hybrid DL and Ensemble ML Models for GDM Prediction

Model Type	Example Models	Accuracy
Hybrid Deep Learning	GRU+LSTM	Up to 99%
Ensemble ML	MajorityVoting, Bagging, Boosting (XGBoost+KNN+RF)	~86%

V. PERFORMANCE METRICS AND COMPARATIVE ANALYSIS

Assessing the effectiveness of machine learning (ML) and deep learning (DL) models in predicting Gestational Diabetes Mellitus (GDM) requires the use of well-established performance evaluation metrics that ensure objectivity and comparability across different studies [1, 2]. These standardised metrics serve as benchmarks to measure how effectively a model can classify patients into high-risk or low-risk categories. In the context of healthcare, particularly with conditions such as GDM that have serious maternal and neonatal implications, the accuracy of detection is of utmost importance. A single metric is often insufficient to capture the nuances of model performance, especially when dealing with imbalanced clinical datasets where the number of non-GDM cases may significantly outnumber GDM cases. For this reason, it becomes critical to employ a combination of multiple metrics rather than relying solely on overall accuracy. Using several indicators provides a more holistic and reliable evaluation, ensuring that models not only perform well in terms of numerical accuracy but also meet the clinical requirement of correctly identifying patients who are genuinely at risk. This multidimensional evaluation approach is essential for building trust in predictive models and facilitating their integration into real-world healthcare systems [2, 3].

A. Key Performance Indicators (KPIs)

Several metrics are commonly employed to evaluate predictive models in healthcare settings [1, 2]. Accuracy measures the proportion of correct predictions—both positive and negative—relative to the total number of cases. However, in datasets with class imbalance, accuracy can be misleading [3]. Precision evaluates the model's ability to limit false positives, ensuring that individuals identified as high-risk are truly in need of intervention [3]. Recall, also known as sensitivity, measures how well a model detects true positive cases [3]. This is vital in clinical situations because failing to identify high-risk patients can lead to severe health outcomes. The F1-score, calculated as the harmonic mean of precision and recall, provides a balanced measure that accounts for both false positives and false negatives [3]. Collectively, these metrics give a thorough understanding of model performance and its potential clinical relevance [1, 3].

B. Comparative Performance across Models

Different machine learning (ML) and deep learning (DL) models exhibit unique advantages and limitations, largely shaped by their architecture, learning capacity, and computational requirements. Table 2 below provides a comparative overview of widely used models in Gestational Diabetes Mellitus (GDM) prediction, evaluating them based on accuracy, precision, recall, F1-score, and computational cost. These metrics collectively highlight the trade-offs involved in selecting the most appropriate model for clinical applications [9, 11].

Model	Accuracy	Precision	Recall	F1-Score	Computational Cost
GRU+LSTM	99%	0.998	0.998	0.998	High
RNN	96%	0.95	0.96	0.955	Medium-High

GRU	90%	0.91	0.90	0.905	Medium
LSTM	82%	0.83	0.82	0.825	Medium-High
Random Forest	85%	0.86	0.85	0.855	Low-Medium
XGBoost	86%	0.87	0.86	0.865	Low-Medium

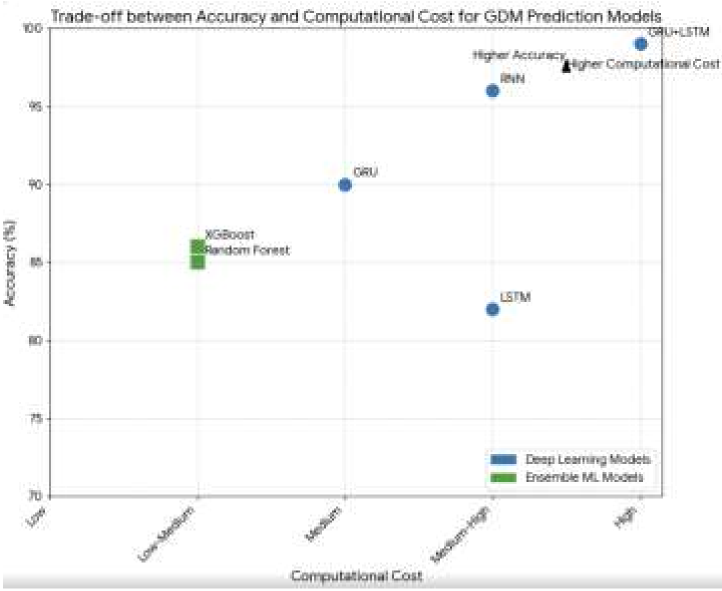
Trade-offs and Insights:

Hybrid deep learning models, particularly GRU+LSTM architectures, demonstrate the highest predictive performance across all metrics [9]. Their ability to capture both short-term and long-term dependencies in sequential patient data makes them exceptionally powerful in clinical prediction tasks [9]. Achieving up to 99% accuracy with near-perfect precision and recall, these models are especially valuable for reducing both false positives and false negatives, thereby minimising risks to patients [9]. However, this superior performance comes at a cost—GRU+LSTM models require high computational resources, longer training times, and advanced hardware such as GPUs or TPUs [9]. This makes them less feasible in low-resource healthcare environments or scenarios where rapid deployment is necessary.

In contrast, Recurrent Neural Networks (RNNs) achieve relatively high performance (96% accuracy) but are more computationally efficient than hybrid models [9]. They remain a viable option when resources are somewhat constrained but not extremely limited. Standalone GRU models also offer a good balance between accuracy (90%) and computational cost, making them useful for institutions that want moderately high performance without the heavy resource burden of GRU+LSTM combinations [9]. On the other hand, LSTM models show lower performance (82% accuracy), primarily due to their higher complexity and susceptibility to overfitting when training data is limited [9].

Amongst the traditional ensemble methods, Random Forest and XGBoost stand out for their balance between predictive performance and efficiency [8, 11]. While their accuracy (85–86%) is lower compared to DL models, their relatively low computational cost and faster training times make them attractive for real-time applications [6, 8, 11]. For example, in settings where quick decisions are critical—such as community clinics or early screening programmes—these models can provide sufficiently accurate predictions without demanding advanced infrastructure. Furthermore, ensemble approaches are easier to interpret and explain to clinicians, enhancing their acceptance in medical practice [6, 9].

The comparison indicates that hybrid deep learning models are most suitable for advanced healthcare institutions with access to powerful computational resources, where maximising predictive accuracy is the top priority [9]. In contrast, ensemble ML models are better suited for practical and scalable deployment, particularly in resource-limited settings [6, 11]. Thus, selecting an appropriate model for GDM prediction is not solely about accuracy—it involves balancing predictive performance with computational efficiency, ease of deployment, and clinical applicability.



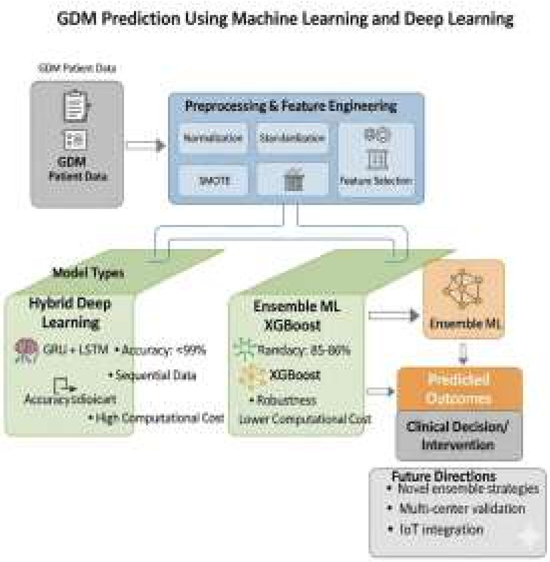
Ultimately, this comparative analysis highlights a spectrum of choices: from resource-intensive but highly accurate hybrid DL models to lightweight, efficient, and moderately accurate ensemble ML methods. The future of GDM prediction may lie in developing adaptive systems that can dynamically switch between these approaches based on available resources and clinical requirements, thereby offering both reliability and accessibility in diverse healthcare settings.

VI. Conclusion and Future Directions

This review demonstrates that machine learning (ML) and deep learning (DL) techniques offer great potential for improving the early detection of GDM [1, 2, 3]. Research indicates that hybrid deep learning architectures, particularly GRU+LSTM models, consistently demonstrate superior predictive performance, with accuracies nearing 99% [9]. By combining the computational efficiency of Gated Recurrent Units (GRU) with the long-term memory capabilities of Long Short-Term Memory (LSTM) networks, these hybrid models can effectively capture complex temporal patterns in sequential clinical data, including glucose readings, insulin fluctuations, and blood pressure trends [9]. This strength makes them highly effective for early identification of high-risk pregnancies, enabling timely medical interventions that can reduce maternal and neonatal complications.

Despite their high accuracy, the practical implementation of hybrid DL models is often limited by substantial computational demands, which may pose challenges in low-resource or real-time clinical environments [9]. Conversely, traditional machine learning approaches, such as Random Forest and XGBoost, offer reliable predictive performance while requiring considerably less computational power, making them more suitable for deployment in resource-limited settings [6, 8, 11]. This presents a critical balance between achieving the highest predictive accuracy and maintaining operational feasibility, which must be carefully considered when selecting models for clinical GDM prediction.

Future research should aim to use larger and more diverse datasets to enhance model generalisability. It should also incorporate explainable AI (XAI) to make models more transparent for clinicians and optimise hybrid and ensemble models for practical deployment in real-world healthcare settings [9, 15]. Such advancements will further enhance the clinical applicability of predictive models, ensuring timely and effective management of GDM.



Gaps in Current Research

Several limitations persist in current GDM prediction research:

- **Limited dataset diversity:** Most studies rely on relatively small, homogeneous datasets, which may reduce the generalisability of models across different populations and ethnic groups [1, 2, 4, 6].
- **Lack of explainable AI (XAI) adoption:** While high accuracy is valuable, clinical adoption requires transparent models that provide interpretable insights to healthcare practitioners [9].
- **Scarce real-time deployment studies:** Few models have been tested in live healthcare environments, limiting understanding of their operational performance and integration challenges [15].

Future Directions

To address these gaps and further enhance GDM prediction, several promising avenues can be explored:

- **Novel ensemble strategies** that combine deep learning and traditional ML approaches, aiming to balance high predictive accuracy with computational efficiency [12, 13, 14].
- **Validation on multi-centre, diverse cohorts** to ensure model generalisability and reliability across different demographic and clinical settings [1, 2, 4].
- **Integration with IoT-enabled health monitoring systems**, allowing continuous real-time data collection, early detection, and timely interventions to improve maternal and fatal outcomes [7, 15].

In summary, while current ML and DL approaches offer powerful tools for GDM prediction, future research should focus on creating interpretable, generalizable, and resource-efficient models that can be seamlessly integrated into clinical workflows for broader impact.

REFERENCES

[1] Z. ZHANG et al.: Machine learning prediction models for gestational diabetes mellitus: meta-analysis. J Med Internet Res, 24 (3), e26634 (2022).

[2] Y. T. WU et al.: Early prediction of gestational diabetes mellitus in the Chinese population via advanced machine learning. J Clin Endocrinol Metab, 106 (3), e1191-e1205 (2021).

[3] G. CUBILLOS et al.: Development of machine learning models to predict gestational diabetes risk in the first half of pregnancy. BMC Pregnancy Childbirth, 23 (1), 469 (2023).

[4] Y. XIONG et al.: Prediction of gestational diabetes mellitus in the first 19 weeks of pregnancy using machine learning techniques. J Matern Fetal Neonatal Med, 35 (13), 2457-2463 (2022).

[5] S. M. LEE et al.: Nonalcoholic fatty liver disease and early prediction of gestational diabetes mellitus using machine learning methods. Clin Mol Hepatol, 28 (1), 105 (2021).

[6] Y. KAYA et al.: The early prediction of gestational diabetes mellitus by machine learning models. BMC Pregnancy Childbirth, 24 (1), 574 (2024).

[7] A. PUSTOZEROV et al.: Machine learning approach for postprandial blood glucose prediction in gestational diabetes mellitus. IEEE Access, 8, 219308-219321 (2020).

[8] GNANADASS: Prediction of gestational diabetes by machine learning algorithms. IEEE Potentials, 39 (6), 32-37 (2020).

[9] B. KURT et al.: Prediction of gestational diabetes using deep learning and Bayesian optimization and traditional machine learning techniques. Med Biol Eng Comput, 61 (7), 1649-1660 (2023).

[10] M. ZHAO et al.: Predictive value of machine learning for the progression of gestational diabetes mellitus to type 2 diabetes: a systematic review and meta-analysis. BMC Med Inform Decis Mak, 25 (1), 18 (2025).

[11] X. HU et al.: Prediction model for gestational diabetes mellitus using the XG Boost machine learning algorithm. Front Endocrinol, 14, 1105062 (2023).

[12] M. KANG et al.: Machine Learning–Based Prediction of Large-for-Gestational-Age Infants in Mothers With Gestational Diabetes Mellitus. J Clin Endocrinol Metab, 110 (5), e1631-e1639 (2025).

[13] M. KUMAR et al.: Population-centric risk prediction modeling for gestational diabetes mellitus: A machine learning approach. Diabetes Res Clin Pract, 185, 109237 (2022).

[14] S. K. BIGDELI et al.: 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 Med Inform Decis Mak, 25 (1), 3 (2025).

[15] H. ZAKY et al.: Machine learning based model for the early detection of Gestational Diabetes Mellitus. BMC Med Inform Decis Mak, 25 (1), 130 (2025).

[16] Y. T. WU et al.: Early prediction of high risk gestational diabetes mellitus via machine learning models. medRxiv, 2020-03 (2020).