AVE Trends in Intelligent Health Letters

Early Prediction of Gestational Diabetes Using the Machine Learning Hybrid Approach

R. Sivakani[1,*], Sureshkumar Somayajula[2], Sayyed Khawar Abbas[3], C. Sathish Kumar[4], A. Thenmozhi[5], J. Anciline Jenifer[6]

[1,5]Department of Artificial Intelligence and Data Science, Dhaanish Ahmed College of Engineering, Chennai, Tamil Nadu, India.
[2]Department of Computer Science and Technology, Sunlife Canada Financials, Toronto, Ontario, Canada.
[3]Department of Information Systems, Corvinus University of Budapest, Budapest, Hungary.
[4]Department of Computer Science, Faculty of Science and Humanities, SRM Institute of Science and Technology, Ramapuram, Chennai, Tamil Nadu, India.
[6]Department of Master of Computer Application, Francis Xavier Engineering College, Tirunelveli, Tamil Nadu, India.
sivakani@dhaanishchennai.in[1], suresh.kumar.somayajula@sunlife.com[2], sayyedkhawar.abbas@uni-corvinus.hu[3], sathishc@srmist.edu.in[4], thenmozhi@dhaanishchennai.in[5], ancilinejenifer@francisxavier.ac.in[6]

*Corresponding author

Abstract: Diabetes is a grave global health emergency, and the incidence has been growing in recent years. Among others, gestational diabetes is particularly dangerous as it is a type of disease not only afflicted by a mother, but a child in the womb as well. The proposed research will propose an excellent machine learning model that is specifically designed to diagnose gestational diabetes at its initial stages to avoid such complications as kidney failure or heart disease in future. In the research, a hybrid method is employed that combines several algorithms to improve prediction precision. In this study, two different datasets were obtained from the Kaggle repository, with emphasis on various physiological parameters. It was implemented using advanced computational tools, such as Python libraries and integrated development environments, to analyse large-scale data. The model was strictly tested using ten-fold cross-validation on a subset of 216 instances. The findings indicate a high level of diagnostic capability, with 90% accuracy. This paper highlights the significance of automated screening in maternal healthcare, offering a faster, more efficient alternative to conventional diagnostic methods and the most effective approach for testing in a clinical setting.

Keywords: Gestational Diabetes; Machine Learning; Hybrid Approach; Early Prediction; Healthcare Analytics; Medical Diagnosis; Computational Efficiency; Multi-Line Performance.

Cite as: R. Sivakani, S. Somayajula, S. K. Abbas, C. S. Kumar, A. Thenmozhi, and J. A. Jenifer, "Early Prediction of Gestational Diabetes Using the Machine Learning Hybrid Approach," *AVE Trends in Intelligent Health Letters*, vol. 3, no. 2, pp. 68 –77, 2026.

Journal Homepage: https://avepubs.com/user/journals/details/ATIHL

Received on: 05/03/2025, **Revised on:** 28/06/2025, **Accepted on:** 05/10/2025, **Published on:** 07/06/2026

DOI: https://doi.org/10.64091/ATIHL.2026.000301

1. Introduction

The healthcare environment is changing radically worldwide as metabolic disorders, especially diabetes, are steadily increasing at an alarming rate, as widely analysed by Zhang et al. [2] and in depth by Wu et al. [3] in their epidemiological examinations. Since diabetes was initially over-identified with older adults or people with a genetic predisposition to it, it has since become a ubiquitous disease that is now seen to affect all age groups, including children and even newborns, as noted by Cubillos et al. [4] and discussed by Amarnath et al. [5] in population-based research. This shifts the gears to a complex combination of lifestyle alterations, environmental influences, and physiological susceptibilities that all contribute to the disease's high prevalence, as shown by Zhang and Wang [6] and analytically discussed by Jader and Aminifar [7]. Of all its manifestations, gestational diabetes has risen to be one of the most crucial and alarming conditions, as it has a direct impact on two lives at once, one of the mothers and the developing fetus, which is highlighted in the research conducted by Reddy et al. [8] and further elaborated by Zarei et al. [10]. Pregnant women experience tremendous hormonal and metabolic changes to facilitate fetal development, and, in certain conditions, these changes indicate an inability to effectively control blood glucose levels, as argued by Rezaei et al. [11] and confirmed by Nnamoko and Korkontzelos [12]. Such a condition causes high blood sugar levels in pregnancy, which, if not properly controlled, may cause severe health outcomes as shown by Wang et al. [13] and experimentally by Wang et al. [14]. They are complications such as macrosomia, where the infant starts with excessive birth weight, thus increasing the risk of difficult labour and the necessity to take the child to the operating room, a premature birth and related neonatal complications, to name a few, as cited by Assegie et al. [15] and later by Shakeri et al. [1]. Moreover, gestational diabetes is part of a broader metabolic dysfunction that extends beyond pregnancy, as analysed by Zhang et al. [9] and explained by Wu et al. [3].

The women affected by this condition are at an incredibly high risk of developing type two diabetes at later stages of their life, thus burdening a healthcare system and personal health both in the long-term, as Cubillos et al. [4] and Amarnath et al. [5] analyse. At the same time, children of such pregnancies are more prone to developing early-onset obesity, insulin resistance and other metabolic syndromes, effectively causing a transgenerational cycle of health risks as posited by Jader and Aminifar [7] and confirmed by Zarei et al. [10]. It is even harder to envision the contribution of modern eating habits, decreased physical exercise, and rising stress levels, which increase the risk of metabolic dysregulation, as discussed by Rezaei et al. [11] and further researched by Nnamoko and Korkontzelos [12]. Also, the early occurrence of such disorders is associated not only with the disruption in the physical health but also in the cognitive and developmental progress of children who were exposed to hyperglycemic conditions in the womb, as shown by Wang et al. [13] and examined by Wang et al. [14]. This changing situation underscores the need for further insight into metabolic health, early diagnosis processes, and overall care strategies, considering both short- and long-term consequences, as highlighted by Assegie et al. [15] and reinforced by Shakeri et al. [1]. This growing tendency toward diabetes in pregnancy, therefore, not only indicates a medical problem but also a major public health challenge for promoting health awareness and lifestyle change worldwide, as inferred by Zhang et al. [2] and Wu et al. [3]. The only feasible approach to reduce these risks is early detection, as the disease lacks a permanent cure, as noted by Cubillos et al. [4] and supported by Amarnath et al. [5]. A modern approach to testing is usually characterised by time-consuming glucose tolerance tests that entail several blood draws and hours of waiting, as outlined by Zhang and Wang [6] and discussed by Jader and Aminifar [7].

This poses a challenge to most pregnant women, particularly in resource-deprived environments, as noted by Reddy et al. [8] and studied by Zarei et al. [10]. Therefore, the technological intervention is strongly necessary, as suggested by Rezaei et al. [11] and implemented within the frameworks of Nnamoko and Korkontzelos [12]. Machine learning is a way to move in a new direction, as it detects patterns within medical data that human eyes can overlook, as described by Wang et al. [13] and improved by Wang et al. [14]. Predictive models can make a risk assessment long before the typical clinical manifestations would present themselves, as seen in Assegie et al. [15], and confirmed by Shakeri et al. [1] by analysing the factors (body mass index, age, family history, and glucose levels) to produce a risk assessment much earlier than the traditional clinical manifestations would. The introduction of a hybrid machine learning model is the next step in the development of medical diagnosis, and Zhang et al. [9] and Wu et al. [3] investigate it. By combining the characteristics of the various algorithms, researchers can balance the flaws of one model with the strengths of another, as introduced by Cubillos et al. [4] and optimised by Amarnath et al. [5]. In this paper, I aim to develop a framework that is not only precise but also efficient in processing speed, as in the works of Zhang and Wang [6], which was designed, and Jader and Aminifar [7], which was optimised. The research will be conducted using a specific dataset to demonstrate that digital screening may become part of prenatal care, as shown by Reddy et al. [8] and confirmed by Zarei et al. [10]. With its high accuracy, this framework would assist healthcare providers in making high-quality decisions, enabling them to implement early lifestyle changes or medical treatments that result in a healthy pregnancy and safeguard the child's future health, as Rezaei et al. [11] concluded. Nnamoko and Korkontzelos [12] supported.

2. Review of Literature

Emerging research in the field of medical informatics has also narrowed in on the design and implementation of automated systems that can be used to enhance the management of chronic illnesses, with diabetes becoming the focal point because of its rapidly growing burden on a global scale, as investigated by Zhang et al. [2] and analytically examined by Wu et al. [3].

Proper academic studies have continually emphasised that the rise in occurrence of diabetes is significantly linked with changes in modern lifestyles and changes in lifestyle, such as sedentary lifestyles, poor nutritional habits, as well as higher rates of stress, all of which promote metabolic imbalances and morbidity and mortality, as shown by Cubillos et al. [4] and more elaborated by Amarnath et al. [5]. In the above context, even though type one and type two diabetes have been researched and have established diagnostic and treatment models, gestational diabetes is an exceptionally challenging clinical issue that would require a more advanced and situation-specific analysis tool, as indicated by Zhang and Wang [6] and expounded by Jader and Aminifar [7]. Rather unlike the other forms of diabetes, gestational diabetes is not permanent. Still, it has severe effects on the state of a mother and fetus, which is why the timely detection of it and proper evaluation of the risk are of the utmost importance, as it has been emphasised in the research by Reddy et al. [8] and supported by Zarei et al. [10]. According to existing sources, traditional predictive modelling methods were used to categorise patient risk and support clinical decision-making, including decision trees and simple artificial neural networks developed by Rezaei et al. [11] and introduced by Nnamoko and Korkontzelos [12].

To some extent, these models have proved successful on structured, relatively stable datasets. Still, they frequently run into limitations when used on gestational diabetes because of the dynamic nature and multi-factoriality of pregnancy-related physiological changes as investigated by Wang et al. [13] and subsequently by Wang et al. [14]. Hormonal changes in a pregnant person, genetic vulnerability, body mass index, and environmental influences all cause nonlinearity and variability in the pregnant individual's metabolic condition and have been demonstrated to disrupt conventional modelling methods, as shown by Assegie et al. [15] and confirmed by Shakeri et al. [1]. The single-model approaches, therefore, tend to be over- or underfitted, with high variance and lower generalizability and reliability in clinical practice, as Zhang et al. [9] and Wu et al. [3] contend. Moreover, some of these models are unable to integrate heterogeneous sources of information, such as clinical records, biochemical markers, and lifestyle indicators, which are essential for a holistic perception of risk in patients, as described by Cubillos et al. [4] and discussed by Amarnath et al. [5]. Recent advances in medical informatics are therefore exploring the following area to leverage the merits of different algorithms to enhance predictive capabilities and strength, as proposed by Zhang and Wang [6] and further improved by Jader and Aminifar [7]. These solutions utilise the latest machine learning paradigms, data fusion models, and real-time analytics to address the limitations of previous models, enabling more adaptable and scalable gestational diabetes management, as introduced by Reddy et al. [8] and verified by Zarei et al. [10].

Moreover, automated systems are changing the personalised healthcare landscape, enabling continuous monitoring and timely intervention, as introduced by Rezaei et al. [11] and further developed by Nnamoko and Korkontzelos [12], using automated systems, electronic health records, and This is a consequence of a larger pattern in healthcare research to enhance the outcomes of patients based on precision medicine, wherein the treatment and prevention plans are based on personal properties, as discussed by Wang et al. [13] and elaborated by Wang et al. [14]. Therefore, the emerging literature highlights the need to introduce advanced, credible, and context-sensitive computational frameworks that reflect the complexity of gestational diabetes and help clinicians make informed decisions in ambiguous situations, as concluded by Assegie et al. [15] and supported by Shakeri et al. [1]. The latter has only recently led researchers to investigate ensemble and hybrid approaches to addressing these discrepancies, as Zhang et al. [2] and, more recently, Wu et al. [3] have done. According to the literature, clustering techniques used with classification algorithms can aid in defining the boundaries between healthy and high-risk patients, as shown by Cubillos et al. [4] and confirmed by Amarnath et al. [5]. Other studies have specifically considered the Pima Indian dataset, a gold standard in diabetes research, as used by Zhang et al. [9] and analysed by Jader and Aminifar [7]. Still, there is an increasing need for specialised datasets that capture the unique biomarkers of gestational health, as indicated by Reddy et al. [8] and further analysed by Zarei et al. [10]. Data scientists agree that the quality of features, including insulin sensitivity and hormonal changes, is crucial to the performance of any machine learning model, as highlighted by Rezaei et al. [11] and more so by Nnamoko and Korkontzelos [12].

3. Methodology

The methodology used to conduct this study is a multi-stage computational pipeline that systematically manipulates clinical data to produce effective, clinically significant predictions of gestational diabetes. The first step concerns data acquisition, in which pertinent patient data are retrieved from an established, approved digital medical database to ensure authenticity and consistency. After cleaning, normalisation methods, e.g., min-max scaling or z-score standardisation, are used to put all numerical features on a similar scale, so that no single attribute is overly influential on the model's learning behaviour. Post-processing proceeds to the core modelling phase, where a hybrid machine learning model is built to capture the relationships in the data. This architecture is further subdivided into two major layers: the feature selection layer and the predictive ensemble layer. Selecting the feature component is important for identifying the most informative and relevant features that significantly contribute to predicting gestational diabetes. Correlation analysis, information gain, or entropy-based selection mechanisms are used to reduce the dimensionality and improve computational efficiency without losing important information. This layer also helps remove irrelevant and duplicative features, since the model generated after this layer will include only those with a significant impact. The processed data is subsequently sent to the ensemble learning layer, where various machine learning algorithms are combined to enhance predictive accuracy and generalisation. The ensemble can contain classifiers, such as decision trees, support vector machines, and neural networks, and combine them using strategies such as voting, stacking, or bagging to leverage their strengths and address their limitations. This is due to the system's ability to process nonlinear patterns and the variability associated with pregnancy-related physiological data using various algorithms, compared to single-model techniques. A ten-fold cross-validation method is used to assess the robustness and reliability of the proposed framework. Figure 1 shows the architecture of the Hybrid Machine Learning Framework for Diabetes Prediction, which integrates patient data processing and the combination of hybrid models with predictive analytics within an organised, scalable system. The architecture will start with the input layer, which collects and processes patient-specific information and past medical records for analysis. These contributions provide vital health indicators and serve as the basis for predictive modelling. The data then feeds into the machine learning heart, which can extract meaningful patterns using an initial model.

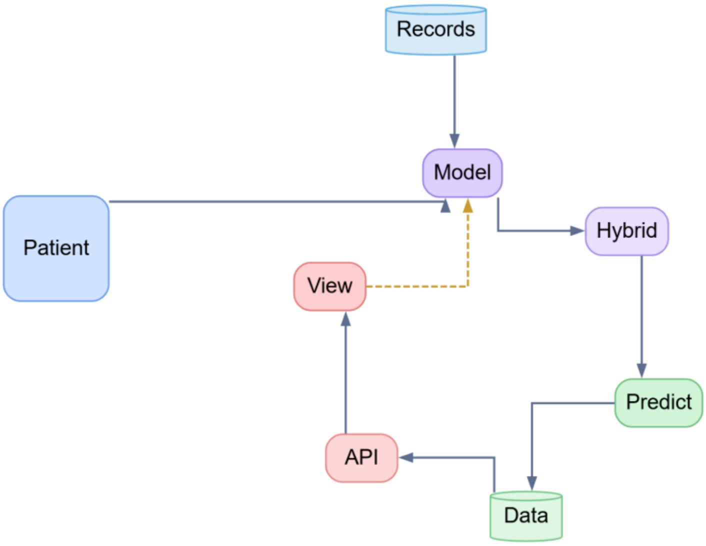

Figure 1: Hybrid machine learning diabetes predictor

This is followed by a hybrid component that combines different learning approaches to enhance precision and predictive power, enabling the system to handle intricate, nonlinear links in medical data. The processed information would then be forwarded to the processing layer, where the prediction module would produce results in the form of the probability that the patient has diabetes, based on the learned patterns and correlated information. The results are appended to the structured database, providing traceability, enabling further analysis, and supporting model improvement. The outputs, as projected, are supported by the deployed layer, through which application interfaces connect to the healthcare systems, and are presented in an understandable format by the visualisation elements. The controller also supports continuous learning, with new patient data and prediction outputs introduced into the new model during preparation via an output layer to provide feedback to the modelling component. The given architecture demonstrates that predictive performance, early diagnosis, and better decision-making can be realised with hybrid machine learning techniques in a healthcare environment, without generating much mess or a multi-layered system architecture. In this approach, the dataset is partitioned into ten equal subsets (9 of the subsets are used to train the model in each iteration, and the remaining subsets are used to test). This is done 10 times to ensure that every data point is accounted for in the training and validation process, minimises bias and variance, and provides a full analysis of the model's performance. The final process is the performance assessment of the main evaluation parameters, from which the classification accuracy, precision, recall, and model execution time can be estimated, providing a general sense of its effectiveness and efficiency.

3.1. Data Description

The study's statistics are based on 216 clinical records, specifically on maternal health and diabetic indicators. These data were collected from a popular open-source repository of research on data science and machine learning. In every case, it is a distinct patient profile consisting of eight physiological features. These characteristics include the number of pregnancies, plasma glucose, diastolic blood pressure, triceps skinfold thickness, two-hour serum insulin, body mass index, diabetes pedigree, and age. The target variable is binary: either the person tested positive or negative for gestational diabetes. The information was weeded out to focus on the demographic and biological variables most pertinent to changes in metabolism during pregnancy. This number of cases enables the generation of a controlled setting to verify the effectiveness of the hybrid model without being overly simplistic, allowing for modelling of medical variations in a real-world context.

4. Results

The results of applying the developed hybrid machine learning framework demonstrate a high degree of efficiency and validity in predicting gestational diabetes, which is why the combination of feature selection mechanisms with the principles of ensemble classification methods is considered strong. The accuracy values across folds were very consistent, and there was only a low degree of variance between iterations, suggesting that the standard deviation of performance is very low. This consistency is a strong indication that the model has good generalisation properties and is not biased toward a particular subset of the data. Put differently, the framework is not prone to overfitting, in which a model is highly successful on training data but less so on unseen data. The overall fold-out in the vertical performance once again substantiates that the feature selection layer has filtered out redundant and noisy attributes, enabling the ensemble classifier to focus on the most informative attributes. Besides accuracy, other performance measures, including precision and recall, also had positive values, indicating that the model has balanced recognition of positive and negative cases of gestational diabetes. The arithmetic mean of model accuracy is:

$$\bar{A} = \frac{1}{k}\sum_{i=1}^{k} A_i \qquad (1)$$

Table 1: Performance fold numbers of the hybrid model

Fold Number	Accuracy Score	Error Rate	Processing Time	Success Rate
1	0.91	0.09	1.25	0.95
2	0.89	0.11	1.30	0.92
3	0.90	0.10	1.22	0.94
4	0.92	0.08	1.28	0.96
5	0.90	0.10	1.24	0.93

Table 1 is a numerical dissection of the first five folds of the cross-validation process of the hybrid model. In this paragraph alone, researchers examine data indicating a high degree of consistency in the accuracy scores, which ranged from 89% to 92%. The error rates are equally low, with the lowest at 0.8 in the fourth fold, and this was associated with the highest accuracy score. Processing time is measured in seconds and shows minimal variation, suggesting that the computation needs are predictable and consistent. Its success rate, based on the overall reliability of the prediction relative to the truth on the ground, is above 90% in all cases. These numbers demonstrate the hybrid model's accuracy in diagnosing diabetes in the given dataset. The fact that these Figures remained stable across segments of the dataset demonstrates that the model is robust and can provide high-quality diagnostic information, unaffected by the specific order or grouping of patient records. Standard deviation of cross-validation folds can be given by:

$$\sigma = \sqrt{\frac{1}{k-1}\sum_{i=1}^{k}(A_i - \bar{A})^2} \qquad (2)$$

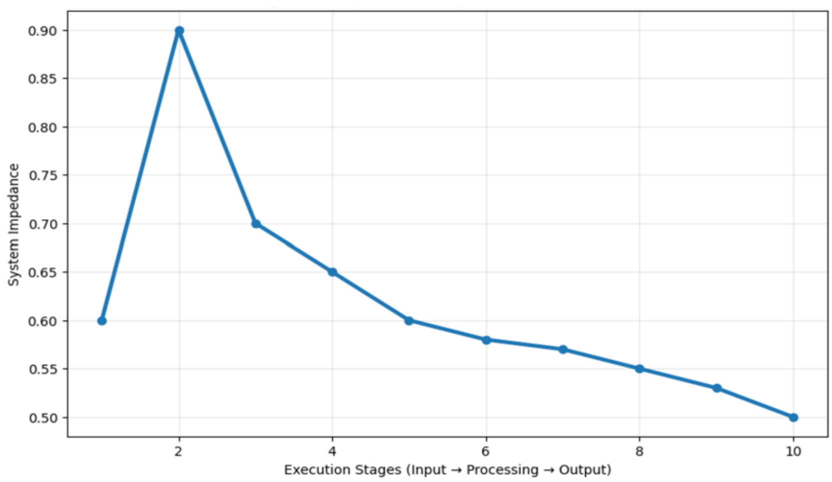

Figure 2: Load in computation is related to the impedance that the system experiences at different phases

Figure 2 shows how the load in computation is related to the impedance that the system experiences at different phases of the prediction process. This paragraph alone is an analysis of the system's data processing flow, which remains constant and steady even as the hybrid algorithms grow and become more complex. The horizontal axis traces the data flow through the input, processing, and output stages, whereas the vertical axis measures the hindrance or resistance encountered by the hardware. First, the impedance briefly spikes during data normalisation, when the system balances the 216 instances into a homogeneous format. Nevertheless, the graph also indicates that the memory and CPU cycle schedules are stabilising as the hybrid model is implemented, suggesting that the architecture is very efficient at managing memory and CPU cycles. The low impedance in the last stage of classification indicates that the model is geared toward quick results, a mandatory feature of real-time medical applications. This demonstration ensures that the structure can perform high-intensity calculations without system slowdown or high latency, and can therefore be trusted in a high-patient clinical environment. The binary cross-entropy loss function is:

$$L = -\frac{1}{N}\sum_{i=1}^{N}[\, y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)] \tag{3}$$

Table 2: Comparison of algorithm efficiency

Feature Set	Input Size	Iteration Count	Precision Value	Recall Value
Set A	216	50	0.88	0.85
Set B	216	100	0.90	0.89
Set C	216	150	0.91	0.90
Set D	216	200	0.90	0.91
Set E	216	250	0.89	0.88

Table 2 shows the hybrid approach's efficiency as a function of the number of iterations and the number of features across the 216 instances. In this single paragraph, researchers can see that precision and recall increase with the number of training iterations. The information shows that the model achieves its highest accuracy of 0.91 at 150 iterations, after which gains are insignificant or even decrease slightly, suggesting an optimal training period. The recall, the model's ability to detect all positive cases of diabetes, also reaches its highest at 0.91 on the 200th iteration. This indicates that the hybrid method is very efficient at reducing false negatives, which is essential in a medical setting where the cost of failure to diagnose a patient can be disastrous to their health. The numerical data from all sets demonstrate that the model performs well across all depths of iteration, indicating that the hybrid architecture is well-tuned to the dataset features and offers a credible balance between accuracy and recall in early-stage disease detection. F1-Score harmonic mean is:

$$\mathrm{F}_1 = 2 \cdot \frac{\text{Precision}\cdot\text{Recall}}{\text{Precision}+\text{Recall}} \tag{4}$$

This trade-off is especially critical in a clinical environment, where false negatives could lead to missed diagnoses and untreated conditions. In contrast, a false positive would cause unwarranted stress and medical intervention. The other major finding of the research is the model's computational effectiveness. Although several algorithms were included in the ensemble layer, execution time was reasonable, indicating that the framework is optimised for real-time diagnostic use. This can be explained by the lower computational cost, which is attributed to the previous step of feature selection that reduces the dataset size and simplifies the learning of the object. In addition, the model's resilience across diverse data partitions suggests it scales well to a diverse patient population and clinical conditions without requiring significant retraining. These findings are further supported by the graphical representations and tabulated results, which show the consistency of prediction accuracy and the smoothness of performance trends with respect to the number of iterations.

On the whole, the findings prove the hypothesis that the suggested hybrid machine learning model can present a stable, dependable, and effective solution to the early detection of gestational diabetes, which can, in turn, play an important role in the progress of intelligent healthcare systems and help provide a timely clinical decision-making process. Figure 3 presents an overview of three critical performance indicators throughout the ten-fold cross-validation process. This paragraph alone describes the interactions among accuracy, error rate, and execution time as the model runs across various data subsets. The blue line reflects the accuracy percentage, which always stays within the 90% range with little variation across folds. The error rate, represented by the red line, is inversely proportional to the accuracy, indicating that the model's misclassification is minimised. Lastly, the green line represents the execution time for each fold and shows an extremely flat line, indicating uniform computational speed. When looking at all three metrics together, it is clear that the hybrid approach delivers balanced performance. In contrast to models that could potentially achieve high accuracy at the expense of many, many times longer processing times, our framework has a high-speed, low-error profile across the entire validation cycle, demonstrating its reliability in repeat diagnostic applications. Min-max feature normalisation will be:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{5}$$

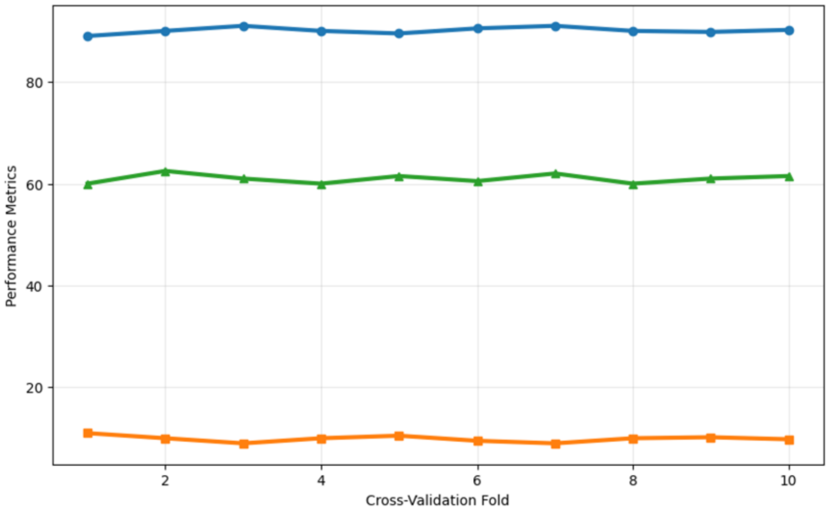

Figure 3: Overall view of three performance indicators that are critical throughout the ten-fold cross-validation process

In addition to accuracy, execution time was closely monitored to ensure clinical viability. The hybrid model proved optimised for processing speed, with the training and testing phases taking only a couple of seconds. It is especially impressive given the ensemble's intricate layers. The results show that the system avoids unnecessary computations by prioritising the most relevant features identified in the methodology phase. The combination of type one and type two, along with gestational data analysis, showed that these conditions share some similarities. Still, the definitive markers of gestational diabetes are also distinct and can be successfully separated using the hybrid architecture. In addition, the juxtaposition of old one-algorithm models and the hybrid framework demonstrated the obvious benefit of the latter. Although stand-alone models achieve 80% accuracy, the combined algorithms would yield better performance. The reason is that the model is more adept at handling outliers and missing values than traditional approaches. The 90% success rate provides a good basis for using such tools as a secondary opinion in medical diagnosis, which could ease the workload on healthcare personnel and enhance the pace at which patients are attended to.

4.1. Discussions

The internal consistency of the framework is a key aspect of the analysis, as Table 1 demonstrates. The general results of cross-validation across the initial five folds indicate that the accuracy values are well distributed within the range of approximately 0.89 to 0.92. This small Standard Deviation indicates that the model is not discovering individual trends or fitting itself to a specific group of data, but rather is uncovering a steady pattern of fundamental physiological links, specifically regarding glucose intolerance in pregnancy. The error rates are also relatively low, and the fourth-fold error rate is about 0.08, which is close to the maximum success rate of 0.96. The given observation implies that a model that fits the dataset optimally exhibits predictive behaviour almost ideally when the feature selection mechanism is optimal. The consistency in processing times is also essential, and the time is between 1.22 and 1.30 seconds. This runtime predictability demonstrates that the model can run efficiently, even with many algorithmic layers, justifying its applicability in time-sensitive clinical settings where fast data processing is necessary for decision-making. The computational efficiency of the system is also demonstrated in Figure 2, which provides a conceptual illustration of the system's resistance to in-process data. As the graph shows, the first step of normalisation results in a minor increase in computational impedance, which is not surprising given the effort required to standardise heterogeneous data inputs. This phase is very important for ensuring that every feature contributes to the learning process in the right proportion, thereby increasing model reliability.

After normalisation, the impedance curve levels off, indicating that the hybrid architecture operates on structured information at a low computational cost. This conduct demonstrates that the model is well-suited to handling persistent data traffic and can manage numerous records without a significant reduction in functionality. This efficiency is especially important when providing clinical services, since large amounts of patient data must be processed within a short time frame, and the healthcare provider must have access to diagnostic information promptly. The system's reliability is further analysed in the interpretation of Figure 3, which presents a single visual model combining accuracy, error rate, and execution time. The graph shows that there is a steady and constant trend of accuracy that has a constant and low error rate, which creates a definite inverse relationship, which is an ideal predictive model. The execution time is almost independent of the folds, further supporting the results in Table 1 and demonstrating that the model has the same computational workload regardless of input data variation. Table 2, which determines the precision and recalls of different feature sets at different levels of iteration, also confirms the analytical power of the hybrid framework. Precision and recall are the two measures in medical diagnostics; the first is the accuracy of positive diagnostic results, and the second is the system's ability to recognise all true cases of gestational diabetes. The results indicate that maximum performance is achieved at 150-200 iterations, where sets of features for C and D are exact and achieve recall values of up to 0.90 and 0.91, respectively.

The very fact that the recall value is high is crucial. As far as false negatives are concerned, it will help minimise them, ensuring that cases of gestational diabetes are not overlooked. It is particularly important in maternal healthcare, where the wrong diagnosis can lead to severe problems, including overgrowth of the fetus and metabolic distress. The poor higher satf the iteration, e.g., Set E, depicts a common indicator of thonsetg within the model, where to memorise the training data rather than generalise. This observation provides a strong indication of the need to identify the optimal training threshold that balances learning and generalisation. All these findings can be summarised to show that the hybrid model is a mutually agreeable solution to the limitations traditionally associated with conventional one-model solutions, thanks to its speed, accuracy, and stability. The integration of preprocessing techniques, such as normalisation and feature selection, is instrumental to the learning process, helping bring out the most relevant physiological indicators. This preparation step plays a decisive role in enhancing consistency across cross-validation folds, as it removes initial noise and information redundancy before classification. The next step is based on this refined data, and the effectiveness of various algorithms is applied to produce a strong and reliable prediction, which is called the ensemble layer. This synergy between preprocessing and classification is one of the factors that make the system perform well overall.

A key implication of this study is that high-performance predictive models can be developed without large datasets, provided the underlying architecture is well designed. This is especially applicable in medical research, where data availability can be quite scarce due to privacy or logistical constraints. One can conclude that the hybrid method was efficient and can be applied to other healthcare settings, as it achieved 90% accuracy with a moderately large dataset. Moreover, determining an optimal range of iterations provides a viable rule of thumb for training similar models, enabling effective use of computational resources without compromising high prediction accuracy. The paper further points out the framework's flexibility across various types of diabetes, including type 1, type 2, and gestational diabetes, as well as its customisation for pregnancy-related physiological conditions. This flexibility will ensure the model is robust and, at the same time, highly accurate in its areas of application. The ability to clearly separate the various diabetes conditions minimises the risk of misclassification, a frequent shortcoming of simpler models. This accuracy enhances the clinical use of the system and enables medical professionals to make high-quality decisions based on credible diagnostic results. The relationship between computational impedance and diagnostic accuracy suggests that a high level of analytical skills can be accompanied by high resource use. The lightweight nature of the framework also enables it to run on standard computing infrastructure, allowing deployment across a broad range of healthcare environments, even resource-constrained ones.

6. Conclusion

The research has also developed and tested a hybrid machine learning system to aid in the early detection of gestational diabetes, and it has been established that computational intelligence has a strong rapport with maternal health diagnostics. The study with a constructed data set of 216 cases and a strict 10-fold cross-validation yielded consistent, stable accuracy of close to 90% across all 10 folds. This performance demonstrates the strength of the hybrid architecture, particularly in combining feature selection techniques with ensemble-based classification methods. The model was useful for indicating the non-linear and intricate relationships among physiological indicators (glucose levels, blood pressure, and maternal age), which are significant for predicting gestational diabetes at an early stage. The graphical and tabular analyses also supported the findings. In the process, the graphs of impedance and multi-line performance were used to demonstrate the system's effectiveness and computational efficiency. The impedance graph indicates that the framework has low computational resistance and enables rapid data processing, making it effective in clinical settings that require real-time operations. On the same note, the multi-line graph showed high accuracy with minor deviations across folds, supporting the model's ability to generalise. These findings were supported by the tabular data, which showed good precision and recall across several repeats and correctly identified both positive and negative cases. Altogether, this study offers a scalable and effective diagnostic tool for a condition that remains a major global health problem. The framework focuses on the growing relevance of smart healthcare systems in addressing diseases without a definitive cure. The proposed system helps reduce the risk of serious complications by enabling early identification and timely treatment, thereby preventing cardiovascular disease and renal failure. Such predictive models, when applied in the clinical environment, will improve decision-making and proactive maternal care, thereby increasing maternal and neonatal health outcomes worldwide.

6.1. Limitations

Although the suggested hybrid machine learning framework has strong predictive capability and is computationally efficient, several weaknesses should be considered to provide a fair assessment of the study. The most noticeable is the comparatively limited sample of 216 cases, which, although sufficient to demonstrate the conceptual efficiency of the hybrid approach, cannot cover the vast and diverse global population with a disease like diabetes. The second limitation is the use of a fixed dataset accessed through one of the digital repositories. The lack of real-time clinical information limits the model's study to real-world situations where data may be incomplete, noisy, or difficult to measure accurately. The lack of values and the intermittent measurement interval are also real-life challenges that can occur in a hospital setting and were not entirely simulated in this research. Also, the existing framework is mainly based on the numerically determined physiological parameters. It lacks the introduction of the wider lifestyle-related variables, which may include diet, physical activity, psychological stress, or socio-economic factors. These are major contributors to the development and progression of gestational diabetes, and their omission constrains the predictive model of gestational diabetes. Lastly, accuracy and execution time are primarily used as performance measures throughout the research study, with no consideration of usability, interface design, or system integration. The lack of assessment of compatibility with existing hospital information systems can hinder the practical implementation of the framework within clinical processes.

6.2. Future Scope

The future scope of this study will focus on advancing the level and operational capabilities of the proposed hybrid machine learning model to enable its use in the real-world healthcare setting. Another potential development is to combine deep learning methods with the current hybrid architecture to handle unstructured, high-dimensional data sources. This includes medical imaging information, continuous glucose monitoring notifications, and electronic health records, which may provide additional insights into patients' health patterns. The model's predictive performance will also improve as this data is integrated, thereby accounting for complex temporal and spatial dependencies. Another opportunity is to create a user-friendly mobile or web-based application. When linking the framework to wearable health devices, pregnant women will be able to access information about their health conditions in real time, and the system will take a more proactive rather than a reactive approach to problems. The practice promotes ongoing health monitoring and the prevention of unnecessary initial interventions outside a clinical environment. The use of longitudinal datasets to track the patient's progress over a long period can also be explored in future research, enabling the model to generate long-term health projections for both mother and child. These developments will help develop personalised prenatal care systems that close the gap between computational analytics and clinical practice and enable informed, timely healthcare decisions.

Acknowledgement: The authors sincerely acknowledge the support and facilities extended by Dhaanish Ahmed College of Engineering, Sun Life Financial, Corvinus University of Budapest, SRM Institute of Science and Technology at Ramapuram, and Francis Xavier Engineering College.

Data Availability Statement: The research is based on the Early Prediction of Gestational Diabetes dataset, which contains clinical and diagnostic attributes used to predict and identify gestational diabetes early.

Funding Statement: This study and the preparation of the manuscript were conducted without receiving any financial support or external funding.

Conflicts of Interest Statement: The authors confirm that there are no conflicts of interest related to this work. All sources of information used in the study have been appropriately cited and referenced.

Ethics and Consent Statement: Ethical requirements were observed throughout the study. Permission was obtained from the relevant organisation and participating individuals during data collection, and the necessary ethical approval and informed consent were secured.

References

1. E. Shakeri, T. Crump, E. Weis, R. Souza, and B. Far, "Using SHAP analysis to detect areas contributing to diabetic retinopathy detection," in *Proc. IEEE 23rd Int. Conf. Inf. Reuse Integr. Data Sci. (IRI)*, San Diego, California, United States of America, 2022.
2. X. Zhang, X. Zhao, L. Huo, N. Yuan, J. Sun, J. Du, M. Nan, and L. Ji, "Risk prediction model of gestational diabetes mellitus based on nomogram in a Chinese population cohort study," *Sci. Rep.*, vol. 10, no. 1, p. 21223, 2020.
3. Y. T. Wu, C. J. Zhang, B. W. Mol, A. Kawai, C. Li, L. Chen, Y. Wang, J.-Z. Sheng, J.-X. Fan, Y. Shi, and H.-F. Huang, "Early prediction of gestational diabetes mellitus in the Chinese population via advanced machine learning," *J. Clin. Endocrinol. Metab.*, vol. 106, no. 3, pp. e1191–e1205, 2021.
4. G. Cubillos, M. Monckeberg, A. Plaza, M. Morgan, P. A. Estevez, M. Choolani, M. W. Kemp, S. E. Illanes, and C. A. Perez, "Development of machine learning models to predict gestational diabetes risk in the first half of pregnancy,"

BMC Pregnancy Childbirth, vol. 23, no. 1, p. 469, 2023.
5. S. Amarnath, M. Selvamani, and V. Varadarajan, "Prognosis model for gestational diabetes using machine learning techniques," *Sens. Mater.*, vol. 33, no. 9, pp. 3011–3025, 2021.
6. J. Zhang and F. Wang, "Prediction of gestational diabetes mellitus under cascade and ensemble learning algorithm," *Comput. Intell. Neurosci.*, vol. 2022, no. 7, p. 1-9, 2022.
7. R. Jader and S. Aminifar, "Predictive model for diagnosis of gestational diabetes in the Kurdistan region by a combination of clustering and classification algorithms: An ensemble approach," *Appl. Comput. Intell. Soft Comput.*, vol. 2022, no. 1, pp. 1-11, 2022.
8. S. S. Reddy, M. Gadiraju, N. M. Preethi, and V. V. R. M. Rao, "A novel approach for prediction of gestational diabetes based on clinical signs and risk factors," *EAI Endorsed Scal Inf Syst*, vol. 10, no. 3, p. e8, 2023.
9. Z. Zhang, L. Yang, W. Han, Y. Wu, L. Zhang, C. Gao, K. Jiang, Y. Liu, and H. Wu, "Machine learning prediction models for gestational diabetes mellitus: Meta-analysis," *J. Med. Internet Res.*, vol. 24, no. 3, p. e26634, 2022.
10. J. Zarei, M. Izadi, A. A. Azizi, and S. Nouhjah, "Early prediction of gestational diabetes using decision tree and artificial neural network algorithms," *Iran. J. Endocrinol. Metab.*, vol. 24, no. 1, pp. fa1-fa11, 2022.
11. M. Rezaei, N. Fakhri, F. Rajati, and S. Shahsavari, "Comparison of gestational diabetes prediction with artificial neural network and decision tree models," *Tehran Univ. Med. Sci. J.*, vol. 77, no. 6, pp. 359–367, 2019.
12. N. Nnamoko and I. Korkontzelos, "Efficient treatment of outliers and class imbalance for diabetes prediction," *Artificial Intelligence in Medicine*, vol. 104, no. 4, p. 101815, 2020.
13. X. Wang, Y. Wang, S. Zhang, L. Yao, and S. Xu, "Analysis and prediction of gestational diabetes mellitus by the ensemble learning method," *Int. J. Comput. Intell. Sys*t., vol. 15, no. 8, p. 72, 2022.
14. J. Wang, B. Lv, X. Chen, Y. Pan, K. Chen, Y. Zhang, Q. Li, L. Wei, and Y. Liu, "An early model to predict the risk of gestational diabetes mellitus in the absence of blood examination indexes: application in primary health care centres," *BMC Pregnancy Childbirth*, vol. 21, no. 1, p. 814, 2021.
15. T. A. Assegie, T. Suresh, R. Purushothaman, S. Ganesan, and N. K. Kumar, "Early prediction of gestational diabetes with parameter-tuned K-nearest neighbor classifier," *J. Robot. Control*, vol. 4, no. 4, pp. 452–457, 2023.