Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)
Volume 9, Nomor 2, Agustus 2024, pp 1106-1117
Terakreditasi Nomor 204/E/KPT/2022 | ISSN: 2527-5771/EISSN: 2549-7839
https://tunasbangsa.ac.id/ejurnal/index.php/jurasik

Enhancing the Diagnosis of Gestational Diabetes Mellitus: A Comparative Study of Deep Learning and Traditional Machine Learning Models on Imbalanced Datasets

Gregorius Airlangga
Universitas Katolik Indonesia Atma Jaya, Indonesia
Email: gregorius.airlangga@atmajaya.ac.id

Abstract

This study aims to diagnose gestational diabetes mellitus (GDM) using advanced deep learning and traditional machine learning models, focusing on handling imbalanced datasets. The research objective is to develop models that can accurately classify GDM cases, which are often underrepresented in medical datasets. The methodology involves the use of both deep learning models, such as the Advanced Hybrid Model and Voting Model, and traditional machine learning models, including Random Forest, Gradient Boosting, and LightGBM. These models were trained on a balanced dataset achieved through the Synthetic Minority Over-sampling Technique (SMOTE) and evaluated using cross-validation and test accuracy metrics. The results indicated that the deep learning models achieved high cross-validation accuracy but faced challenges in classifying GDM cases on the test set, with lower precision and recall for the minority class. Traditional machine learning models also demonstrated strong performance but similarly struggled with sensitivity towards GDM cases. The study concludes that while these models show promise in diagnosing GDM, further refinement is necessary to improve their ability to handle imbalanced datasets. Future research should explore advanced techniques to enhance the detection of GDM cases, contributing to more reliable and accurate medical diagnostics.

Keywords: *Gestational Diabetes Mellitus (GDM), Deep Learning, Machine Learning, Class Imbalance, Medical Diagnostics*

1. INTRODUCTION

Gestational diabetes mellitus (GDM) is a significant health concern affecting pregnant women worldwide. Characterized by elevated blood glucose levels, GDM poses considerable risks to both mothers and their unborn children [1]–[3]. If not managed properly, GDM can lead to complications such as preterm birth, macrosomia, and an increased likelihood of developing type 2 diabetes later in life [4], [5]. These risks underscore the importance of early detection and effective management of GDM, particularly in regions with limited access to healthcare resources [6]. In such contexts, periodic testing of pregnant women is often the only feasible method for monitoring their health, making the timely and accurate diagnosis of GDM crucial for mitigating its associated risks [7]. Recent advancements in machine learning and artificial intelligence have revolutionized various fields, including healthcare. Intelligent systems designed using machine learning algorithms have demonstrated immense potential in improving diagnostic accuracy, especially in areas with limited medical infrastructure [8]–[10]. These systems offer a promising solution for the early detection and management of GDM, thereby enhancing the quality of maternal and neonatal care [11].

The application of machine learning in healthcare has gained significant traction over the past decade. Numerous studies have explored the use of various algorithms for the early detection and diagnosis of diseases, including diabetes. Traditional models such as logistic regression, support vector machines (SVM), and decision trees have been widely used in medical diagnostics due to their simplicity and interpretability [12]–[14]. However, these models often fall short when dealing with complex datasets, particularly those involving nonlinear relationships and high-dimensional data [15]. As a result, there has been a growing interest in the application of deep learning techniques, which have shown superior performance in handling such challenges [16]. Deep learning, a subset of machine learning, involves the use of neural networks with multiple layers (hence the term "deep") to model complex patterns in data. Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and their variants have been successfully applied in various medical imaging and signal processing tasks. For instance, CNNs have been used to analyze medical images for disease diagnosis, while RNNs have been employed in predicting patient outcomes based on sequential health records. These models have demonstrated remarkable accuracy in tasks that require the extraction of intricate features from large datasets [17]–[19].

In the context of diabetes diagnosis, several studies have explored the use of machine learning models. For instance, a study by [20] provided a comprehensive review of machine learning and data mining approaches applied to diabetes research. The authors highlighted the potential of these techniques in improving the prediction and management of diabetes, emphasizing the need for more advanced models that can handle the complexity of diabetes-related data. Another study by [21] employed a deep learning approach to predict the onset of type 2 diabetes, demonstrating that deep models outperform traditional methods in terms of prediction accuracy.

Despite these advancements, the application of deep learning specifically for GDM diagnosis remains relatively underexplored. Most existing studies have focused on general diabetes prediction or management, with limited attention given to the unique challenges associated with GDM [22]. This gap in the literature highlights the need for research that specifically addresses the diagnostic challenges posed by GDM, particularly in regions with limited healthcare resources. The urgency of developing effective diagnostic tools for GDM cannot be overstated. GDM is associated with significant health risks for both the mother and the fetus, including preeclampsia, cesarean delivery, and neonatal complications. In regions with limited access to healthcare, these risks are exacerbated by the lack of regular monitoring and timely diagnosis. The development of an intelligent diagnostic system that can accurately identify GDM in its early stages would be a major step forward in improving maternal and neonatal health outcomes [23].

Furthermore, the global prevalence of GDM is on the rise, driven by factors such as increasing maternal age, obesity, and changes in lifestyle [24]. This trend underscores the need for scalable and accessible diagnostic solutions that can be implemented in diverse healthcare settings. Machine learning-based systems offer a viable solution, as they can be trained on large datasets and deployed in various healthcare environments, including those with limited resources. The current state of the art in GDM diagnosis primarily involves the use of traditional machine learning models, such as logistic regression, decision trees, and random forests. These models have been widely used due to their simplicity and ease of interpretation. However, they are often limited in their ability to capture complex relationships within the data, particularly in cases where the data is high-dimensional or exhibits non-linear patterns [25].

Deep learning models, on the other hand, have shown significant promise in addressing these limitations. By leveraging the power of neural networks, deep learning models can automatically learn and extract complex features from the data, leading to improved diagnostic accuracy. Recent studies have demonstrated the effectiveness of deep learning in various medical diagnostic tasks, including the detection of diabetic retinopathy and the prediction of cardiovascular disease [26]–[28]. However, the application of these models specifically to GDM diagnosis is still in its early stages [23]. In this study, we aim to advance the state of the art by developing and evaluating deep learning models for the diagnosis of GDM. Our approach involves the use of advanced deep learning techniques, including convolutional layers, dropout regularization, and batch normalization, to build a robust and accurate diagnostic model. We compare the performance of these models with traditional machine learning methods, providing a comprehensive analysis of their relative strengths and weaknesses.

The remainder of this article is organized as follows. In the next section, we provide a detailed overview of the dataset used in this study, including the data collection process and the specific features considered in our analysis. We then describe the methodology employed in developing and evaluating our models, including the preprocessing steps, model architecture, and evaluation metrics. The results of our experiments are presented and discussed in the subsequent section, where we compare the performance of deep learning models with traditional machine learning approaches. Finally, we conclude the article with a discussion of the implications of our findings, the limitations of the study, and potential directions for future research.

2. RESEARCH METHODOLOGY

2.1. Data Preprocessing

The initial phase of our research methodology is dedicated to the comprehensive preprocessing of the dataset, a critical step to ensure the data's suitability for training and evaluating our predictive models. The dataset, sourced from laboratories in the Kurdistan region, comprises various features collected from pregnant women, with both gestational diabetes mellitus (GDM) and non-GDM cases represented. These features include continuous variables, such as blood glucose levels and body mass index, categorical variables like age group and pregnancy history, and binary variables indicating the presence or absence of GDM. To standardize the continuous variables within the dataset, we apply MinMax Scaling, a technique that rescales each feature to a fixed range between 0 and 1. This standardization is particularly important in deep learning models, where unscaled features can lead to problems such as gradient vanishing or explosion, ultimately affecting the model's convergence during training. Mathematically, MinMax Scaling is expressed as presented in the equation 1.

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{1}$$

Where (X') represents the scaled feature, (X) is the original feature value, (X_{min}) the minimum value of the feature across all instances in the dataset, and (X_{max}) the maximum value of the feature across all instances. This transformation ensures that all continuous variables contribute equally during model training, avoiding any one feature from disproportionately influencing the learning process. For instance, a blood glucose level originally ranging from 70 to 200 mg/dL would be rescaled to a range of 0 to 1, with the lowest value mapped to 0 and the highest to 1. Categorical variables in the dataset, such as age group or pregnancy history, must be converted into a numerical format suitable for input into machine learning algorithms. For this purpose, we utilize one-hot encoding, which creates binary indicators for each category within a variable. For example, an age group variable with categories '18-25', '26-35', and '36-45' would be transformed into three separate binary columns, each indicating the presence (1) or absence (0) of a particular age group. Mathematically, if we denote the original categorical variable as (C) with categories ($C_1, C_2, \ldots, C_n$), the one-hot encoded representation would be a vector (V) of length (n), where $V = [v_1, v_2, \ldots, v_n]$ with ($v_i = 1$) if the instance belongs to category (C_i) and ($v_i = 0$) otherwise. This transformation allows the categorical data to be effectively utilized in our models without imposing any ordinal relationships between the categories.

Class imbalance is a prevalent issue in medical datasets, where the number of positive cases (e.g., GDM cases) is often significantly lower than the number of negative cases. This imbalance can lead to biased models that perform well on the majority class but poorly on the minority class. To mitigate this, we employ the Synthetic Minority Over-sampling Technique (SMOTE), which generates synthetic examples for the minority class by interpolating between existing samples. The SMOTE algorithm operates as follows: for each minority class instance (x_i), the algorithm identifies its k-nearest neighbors ($\{x_1, x_2, \ldots, x_k\}$) within the minority class. A random neighbor (x_j) is selected, and a synthetic instance (x_{new}) is generated as $x_{\text{new}} = x_i + \lambda \cdot (x_j - x_i)$, where (λ) is a random number drawn from a uniform distribution in the range [0, 1]. This process generates new samples that lie along the line segment connecting (x_i) and (x_j), effectively expanding the decision region of the minority class. By applying SMOTE, we increase the representation of the minority class in the training dataset, reducing the likelihood of the model being biased towards the majority class. This balanced dataset ensures that the model learns equally from both positive and negative cases, improving its generalization capability.

To evaluate the performance of our models on unseen data, we split the dataset into training and testing sets. The splitting is performed using an 80-20 ratio, where 80% of the data is allocated to training the model and 20% to testing its performance. Crucially, this split is conducted in a stratified manner, meaning that the class distribution (i.e., the proportion of GDM to non-GDM cases) is preserved across both the training and testing sets. Mathematically, the stratified split can be described as follows: given a dataset *(D)* with *(N)* instances and a binary target variable *(y)* representing the presence of GDM (1 for positive cases and 0 for negative cases), we divide *(D)* into a training set (D_{train}) and a testing set (D_{test}) such that $D_{\text{train}} = \{x_i, y_i\}_{i=1}^{0.8N}$, $D_{\text{test}} = \{x_i, y_i\}_{i=0.8N+1}^{N}$, where $(P(y_{\text{train}} = 1) = P(y_{\text{test}} = 1) = P(y = 1))$. This ensures that the model is trained and tested on datasets that are representative of the overall population, providing a realistic assessment of its performance.

In addition to the aforementioned steps, outlier detection is performed to identify and handle anomalous data points that could potentially skew the model's learning process. Outliers are detected using the z-score method, which measures how many standard deviations a data point is from the mean of its corresponding feature. The z-score for a feature value (x_i) is calculated as $z_i = \frac{x_i - \mu}{\sigma}$, where (μ) is the mean and (σ) is the standard deviation of the feature. Data points with a z-score greater than a specified threshold (typically 3 or -3) are considered outliers and are either removed or capped to the threshold value. By applying these preprocessing techniques, we ensure that the dataset is clean, balanced, and standardized, providing an optimal foundation for training our predictive models. Each step is carefully designed to enhance the model's ability to learn from the data and generalize to unseen cases, ultimately improving the accuracy and reliability of our GDM diagnosis system.

2.2. Model Selection

The core of our research methodology lies in the meticulous selection and implementation of both deep learning models and traditional machine learning models for the diagnosis of gestational diabetes mellitus (GDM). This approach is designed to leverage the strengths of advanced deep learning architectures while integrating the robust characteristics of traditional machine learning techniques, ultimately aiming to achieve superior diagnostic accuracy. The selection process involves a careful consideration of the models' ability to handle complex and high-dimensional data, as well as their interpretability and computational efficiency.

2.2.1. Deep Learning Model

Our primary deep learning model is a fully connected feedforward neural network, commonly referred to as a Deep Neural Network (DNN). The architecture of the DNN is crafted to capture intricate nonlinear relationships between the input features and the target variable, which is the diagnosis of GDM. The network comprises multiple layers, including an input layer, several hidden layers, and an output layer. Each hidden layer is responsible for applying a nonlinear transformation to the input data, enabling the network to learn complex patterns within the dataset. The activation function employed in the hidden layers is the Rectified Linear Unit (ReLU), mathematically defined as $f(x) = \max(0, x)$. The

ReLU function introduces nonlinearity into the model, allowing it to approximate complex functions that linear models cannot. ReLU is particularly favored in deep learning because it mitigates the vanishing gradient problem, which can occur during the backpropagation process when training deep networks. By zeroing out negative inputs, ReLU ensures that the gradient does not diminish to zero, thereby maintaining effective gradient flow and facilitating faster convergence.

To prevent overfitting, which is a common challenge in deep learning models, dropout regularization is incorporated into the hidden layers. During each training iteration, a fraction of the neurons (specified by the dropout rate, which is set to 0.5) is randomly set to zero. This process effectively forces the network to rely on a distributed representation of the data rather than becoming overly dependent on any single neuron. Dropout can be mathematically represented as presented in the equation 2.

$$h_{l+1} = f(W_{l+1} \cdot (r_l \odot h_l) + b_{l+1}) \tag{2}$$

where (h_{l+1}) is the output of the layer $(l+1)$, (W_{l+1}) is the weight matrix, (r_l) is a binary mask with entries sampled from a Bernoulli distribution with probability (p), $(\odot)$ denotes element-wise multiplication, (h_l) is the output of the previous layer, and (b_{l+1}) is the bias term. The dropout process ensures that the network generalizes well to new data by reducing the risk of overfitting. The final layer of the DNN employs a sigmoid activation function, which maps the output to a probability value in the range [0,1]. The sigmoid function is mathematically expressed as $\sigma(x) = \frac{1}{1+e^{-x}}$ where (x) is the input to the neuron. The sigmoid function is particularly well-suited for binary classification tasks, as it provides a probabilistic interpretation of the model's predictions. In the context of GDM diagnosis, the output of the sigmoid function represents the probability that a given patient has gestational diabetes.

2.2.2. Traditional Machine Learning Models

In addition to the DNN, several traditional machine learning models are employed for comparison, each selected for its ability to handle different aspects of the dataset and provide insights into the model's performance. The Random Forest Classifier (RFC) is a robust ensemble learning method that constructs a multitude of decision trees during training. Each decision tree is trained on a bootstrap sample of the data, and the final prediction is made by aggregating the predictions of all individual trees. For classification tasks, the final prediction is determined by majority voting, while for regression tasks, it is the average of the predictions. Mathematically, the prediction of a Random Forest can be represented as $\hat{y} = \frac{1}{T}\sum_{t=1}^{T} h_t(x)$ where $(\hat{y})$ is the predicted value, (T) is the total number of trees in the forest, and $(h_t(x))$ is the prediction of the (t)-th tree. The Random Forest model is known for its ability to reduce overfitting by averaging multiple trees, each of which may have a high variance individually but collectively provide a more stable and accurate prediction. The Gradient Boosting Classifier (GBC) is another powerful ensemble technique that builds trees sequentially, where each tree is trained to correct the errors made by the previous ones. The model minimizes a differentiable loss function using gradient descent, allowing it to iteratively improve its performance. The prediction function of GBC is given by $F_m(x) = F_{m-1}(x) + \alpha \cdot h_m(x)$, where $(F_m(x))$ is the prediction at the (m)-th stage, $(F_{m-1}(x))$ is the prediction at the previous stage, $(h_m(x))$ is the new base learner (typically a decision tree) added at stage (m), and (α) is the learning rate, which controls the contribution of each tree to the final prediction. Gradient Boosting is particularly effective in reducing bias, as each new tree focuses on the residuals (errors) of the previous model, gradually improving accuracy.

The Stacking Classifier is a meta-model that combines the predictions of multiple base models, including the DNN, RFC, and GBC. The base models are first trained independently on the dataset, and their predictions are used as inputs to a second-level model, which makes the final prediction. This approach leverages the strengths of different models, combining their outputs to create a more accurate and robust classifier. The stacking process can be mathematically represented as presented in the equation 3.

$$\hat{y} = g\left(h_1(x), h_2(x), \ldots, h_n(x)\right) \quad (3)$$

where $(\hat{y})$ is the final prediction, ($h_{1(x)}, h_{2(x)}, \ldots, h_{n(x))}$ are the predictions of the base models, and $(g(\cdot))$ is the function learned by the second-level model, which could be a simple linear regression or a more complex model like a neural network. By combining the outputs of multiple models, stacking can reduce the risk of overfitting and improve generalization to new data. The selection and integration of these models allow for a comprehensive comparison of deep learning and traditional machine learning approaches. Each model is evaluated based on its ability to accurately diagnose GDM, providing valuable insights into the advantages and limitations of different methodologies in this critical healthcare application. The combined use of these models ensures a robust and reliable diagnostic system, capable of supporting clinical decision-making with high accuracy.

3. RESULTS AND DISCUSSION

The implementation of deep learning models in the diagnosis of gestational diabetes mellitus (GDM) yielded significant improvements in classification performance after addressing class imbalance using SMOTE. Initially, the dataset exhibited a notable imbalance, with 642 non-GDM cases and only 167 GDM cases. After applying SMOTE, the class distribution was balanced, with 642 cases for both classes, ensuring that the models were trained on a more representative dataset. As presented in the table 1, the Advanced Hybrid Model with Deep Learning achieved a cross-validation accuracy of 89.64%, while the Voting Model with Deep Learning followed closely with an accuracy of 89.49%. These results demonstrate the effectiveness of combining multiple deep learning and traditional machine learning models to improve classification accuracy.

When evaluating the models on the test set, the Advanced Hybrid Model with Deep Learning achieved an accuracy of 83.25%. The model showed strong performance in classifying non-GDM cases, with a precision of 0.91 and a recall of 0.87, resulting in an F1-score of 0.89. However, the model's performance was less effective in identifying GDM cases, with a precision of 0.60, a recall of 0.68, and an F1-score of 0.64. The overall macro-averaged F1-score was 0.76, indicating that while the model performed well on the majority class, there is room for improvement in classifying the minority class.

The Voting Model with Deep Learning slightly outperformed the Advanced Hybrid Model in test accuracy, achieving 83.74%. This model also demonstrated strong performance in classifying non-GDM cases, with a precision of 0.91 and a recall of 0.88, resulting in an F1-score of 0.89. For GDM cases, the model achieved a precision of 0.62, a recall of 0.69, and an F1-score of 0.65. The macro-averaged F1-score was 0.77, indicating a slightly better balance between the precision and recall of both classes compared to the Advanced Hybrid Model.

In addition to deep learning models, we evaluated the performance of several traditional machine learning models. The Random Forest Classifier (RFC) emerged as the top performer among these models, achieving a cross-validation accuracy of 90.19%, closely followed by the Blended Model at 90.19% and LightGBM at 89.10%. These results underscore the robustness of ensemble methods, which benefit from the aggregation of multiple models to improve predictive accuracy. The Gradient Boosting Classifier (GBC) also performed well, with a cross-validation accuracy of 86.68%. Other models such as XGBoost, AdaBoost, and Support Vector Machine (SVM) achieved cross-validation accuracies of 88.47%, 84.58%, and 81.31%, respectively, highlighting the efficacy of boosting techniques in handling complex datasets.

When evaluated on the test set, the Random Forest Classifier achieved an accuracy of 81.28%. It demonstrated strong performance in classifying non-GDM cases, with a precision of 0.89, a recall of 0.87, and an F1-score of 0.88. However, similar to the deep learning models, it struggled with GDM cases, achieving a precision of 0.58, a recall of 0.63, and an F1-score of 0.60. The macro-averaged F1-score was 0.74, indicating a need for further refinement in detecting GDM cases. The Gradient Boosting Classifier showed a slightly better test accuracy of 82.27%, with a precision of 0.90, recall of 0.87, and F1-score of 0.88 for non-GDM cases. For GDM cases, it achieved a precision of 0.60, recall of 0.65, and F1-score of 0.63. The macro-averaged F1-score was 0.75, indicating a slight improvement in the balance between the classes compared to the Random Forest Classifier.

The AdaBoost, Support Vector Machine, and Logistic Regression models achieved test accuracies of 79.80%, 79.80%, and 77.34%, respectively. These models consistently showed good performance in classifying non-GDM cases, but their performance on GDM cases was less robust. The XGBoost model achieved a test accuracy of 80.79%, with similar trends observed in precision, recall, and F1-scores. The LightGBM model, while achieving a strong cross-validation accuracy of 89.10%, showed a test accuracy of 81.77%. The model's performance on non-GDM cases was commendable, with a precision of 0.90, recall of 0.87, and F1-score of 0.88. However, its performance on GDM cases was more modest, with a precision of 0.58, recall of 0.64, and F1-score of 0.61. The macro-averaged F1-score for LightGBM was 0.75, indicating room for improvement in handling class imbalance.

The results indicate that deep learning models, particularly the Advanced Hybrid Model and Voting Model with Deep Learning, can achieve high classification accuracy when trained on balanced datasets. However, these models still face challenges in accurately identifying GDM cases, as evidenced by the lower precision, recall, and F1-scores for the minority class. This suggests that while deep learning models are powerful tools for classification tasks, additional techniques such as advanced data augmentation, more sophisticated model architectures, or cost-sensitive learning may be necessary to improve performance on imbalanced datasets. Traditional machine learning models, particularly ensemble methods like Random Forest, Gradient Boosting, and LightGBM, demonstrated competitive performance. These models consistently achieved high accuracy and performed well in classifying non-GDM cases. However, similar to the deep learning models, they struggled with the minority class, indicating a common challenge across model types in handling class imbalance.

Table 1. Classification Results

Model	Cross-Validation Accuracy (%)	Test Accuracy (%)	Precision (Non-GDM)	Recall (Non-GDM)	F1-Score (Non-GDM)	Precision (GDM)	Recall (GDM)	F1-Score (GDM)
Advanced Hybrid Model with Deep Learning	89.64	83.25	0.91	0.87	0.89	0.60	0.68	0.64
Voting Model with Deep Learning	89.49	83.74	0.91	0.88	0.89	0.62	0.69	0.65
Random Forest	90.19	81.28	0.89	0.87	0.88	0.58	0.63	0.60
Gradient Boosting	86.68	82.27	0.90	0.87	0.88	0.60	0.65	0.63
Ada Boost	84.58	79.80	0.84	0.88	0.86	0.66	0.58	0.62
Support Vector Machine	81.31	79.80	0.85	0.88	0.86	0.64	0.58	0.61
Logistic Regression	81.46	77.34	0.82	0.87	0.84	0.64	0.53	0.58
XGBoost	88.47	80.79	0.88	0.87	0.87	0.60	0.61	0.61
LightGBM	89.10	81.77	0.90	0.87	0.88	0.58	0.64	0.61

4. CONCLUSION

The study focused on the application and evaluation of both deep learning and traditional machine learning models for the diagnosis of gestational diabetes mellitus (GDM). The results of this study highlight the strengths and limitations of various models in classifying GDM cases, particularly in the context of imbalanced datasets. The deep learning models, including the Advanced Hybrid Model and Voting Model, demonstrated high accuracy during cross-validation, achieving

89.64% and 89.49%, respectively. However, their performance on the test set, with accuracies of 83.25% and 83.74%, revealed a discrepancy, particularly in detecting GDM cases. While these models excelled in classifying non-GDM cases, their sensitivity towards GDM cases was lower, as indicated by the lower precision and recall scores for the minority class. This suggests that while deep learning models are effective in capturing complex patterns in data, further refinement is needed to improve their ability to generalize well on imbalanced datasets.

Traditional machine learning models, particularly ensemble methods such as Random Forest, Gradient Boosting, and LightGBM, also exhibited strong performance, with the Random Forest achieving a cross-validation accuracy of 90.19%. These models consistently performed well in classifying non-GDM cases but faced challenges similar to the deep learning models in accurately identifying GDM cases. The relatively high accuracy of these models underscores the robustness of ensemble techniques; however, the lower performance metrics for the minority class indicate that addressing class imbalance remains a critical challenge. In conclusion, while both deep learning and traditional machine learning models have shown promise in diagnosing GDM, the study highlights the ongoing challenge of improving model sensitivity to minority classes in imbalanced datasets. Future research should explore advanced techniques such as cost-sensitive learning, data augmentation, and more sophisticated ensemble strategies to enhance the detection of GDM cases. The findings of this study contribute to the growing body of literature on the application of machine learning in healthcare, emphasizing the need for balanced and robust approaches in medical diagnostics.

REFERENCES

[1] A. A. Choudhury and V. D. Rajeswari, “Gestational diabetes mellitus-A metabolic and reproductive disorder,” *Biomed. \& Pharmacother.*, vol. 143, p. 112183, 2021.

[2] D. F. Dłuski, E. Wolińska, and M. Skrzypczak, “Epigenetic changes in gestational diabetes mellitus,” *Int. J. Mol. Sci.*, vol. 22, no. 14, p. 7649, 2021.

[3] S. Sharma, S. Banerjee, P. M. Krueger, and S. M. Blois, “Immunobiology of gestational diabetes mellitus in post-medawar era,” *Front. Immunol.*, vol. 12, p. 758267, 2022.

[4] A. Kouhkan *et al.*, “Gestational diabetes mellitus: Major risk factors and pregnancy-related outcomes: A cohort study,” *Int. J. Reprod. Biomed.*, vol. 19, no. 9, p. 827, 2021.

[5] M. Darbandi *et al.*, “Prevalence of gestational diabetes and its association with stillbirth, preterm birth, macrosomia, abortion and cesarean delivery: a national prevalence study of 11 provinces in Iran,” *J. Prev. Med. Hyg.*, vol. 62, no. 4, p. E885, 2021.

[6] E. U. Alejandro *et al.*, “Gestational diabetes mellitus: a harbinger of the vicious cycle of diabetes,” *Int. J. Mol. Sci.*, vol. 21, no. 14, p. 5003, 2020.

[7] N. Beamish, “Continuous Glucose Monitoring and Diabetes Management During Pregnancy,” 2020.

[8] M. Injadat, A. Moubayed, A. B. Nassif, and A. Shami, “Machine learning towards intelligent systems: applications, challenges, and opportunities,” *Artif. Intell. Rev.*, vol. 54, no. 5, pp. 3299–3348, 2021.

[9] S. Agarwal, A. S. Yadav, V. Dinesh, K. S. S. Vatsav, K. S. S. Prakash, and S. Jaiswal, “By artificial intelligence algorithms and machine learning models to diagnosis cancer,” *Mater. Today Proc.*, vol. 80, pp. 2969–2975, 2023.

[10] K. Kumar, P. Kumar, D. Deb, M.-L. Unguresan, and V. Muresan, "Artificial intelligence and machine learning based intervention in medical infrastructure: a review and future trends," in *Healthcare*, 2023, vol. 11, no. 2, p. 207.
[11] N. Garg, S. K. Arunan, S. Arora, and K. Kaur, "Application of mobile technology for disease and treatment monitoring of gestational diabetes mellitus among pregnant women: a systematic review," *J. Diabetes Sci. Technol.*, vol. 16, no. 2, pp. 491–497, 2022.
[12] P. Shukla, A. Verma, Abhishek, S. Verma, and M. Kumar, "Interpreting SVM for medical images using Quadtree," *Multimed. Tools Appl.*, vol. 79, pp. 29353–29373, 2020.
[13] M. Siddiq, "Use of Machine Learning to predict patient developing a disease or condition for early diagnose," *Int. J. Multidiscip. Sci. Arts*, vol. 1, no. 1, 2022.
[14] T. Chen *et al.*, "A decision tree-initialised neuro-fuzzy approach for clinical decision support," *Artif. Intell. Med.*, vol. 111, p. 101986, 2021.
[15] S. Salerno and Y. Li, "High-dimensional survival analysis: Methods and applications," *Annu. Rev. Stat. its Appl.*, vol. 10, no. 1, pp. 25–49, 2023.
[16] S. Dong, P. Wang, and K. Abbas, "A survey on deep learning and its applications," *Comput. Sci. Rev.*, vol. 40, p. 100379, 2021.
[17] M. M. Islam, S. Nooruddin, F. Karray, and G. Muhammad, "Human activity recognition using tools of convolutional neural networks: A state of the art review, data sets, challenges, and future prospects," *Comput. Biol. Med.*, vol. 149, p. 106060, 2022.
[18] J. A. Nasir, O. S. Khan, and I. Varlamis, "Fake news detection: A hybrid CNN-RNN based deep learning approach," *Int. J. Inf. Manag. Data Insights*, vol. 1, no. 1, p. 100007, 2021.
[19] R. Al-Shabandar, A. Jaddoa, P. Liatsis, and A. J. Hussain, "A deep gated recurrent neural network for petroleum production forecasting," *Mach. Learn. with Appl.*, vol. 3, p. 100013, 2021.
[20] F. A. Khan, K. Zeb, M. Al-Rakhami, A. Derhab, and S. A. C. Bukhari, "Detection and prediction of diabetes using data mining: a comprehensive review," *IEEE Access*, vol. 9, pp. 43711–43735, 2021.
[21] Y. Jian, M. Pasquier, A. Sagahyroon, and F. Aloul, "A machine learning approach to predicting diabetes complications," in *Healthcare*, 2021, vol. 9, no. 12, p. 1712.
[22] R. Modzelewski, M. M. Stefanowicz-Rutkowska, W. Matuszewski, and E. M. Bandurska-Stankiewicz, "Gestational diabetes mellitus—recent literature review," *J. Clin. Med.*, vol. 11, no. 19, p. 5736, 2022.
[23] H. D. McIntyre, A. Kapur, H. Divakar, and M. Hod, "Gestational diabetes mellitus—innovative approach to prediction, diagnosis, management, and prevention of future NCD—mother and offspring," *Front. Endocrinol. (Lausanne).*, vol. 11, p. 614533, 2020.
[24] J. R. Sparks, N. Ghildayal, M.-F. Hivert, and L. M. Redman, "Lifestyle interventions in pregnancy targeting GDM prevention: looking ahead to precision medicine," *Diabetologia*, vol. 65, no. 11, pp. 1814–1824, 2022.
[25] M. M. Khan and W. Yu, "Robotune: high-dimensional configuration tuning for cluster-based data analytics," in *Proceedings of the 50th international conference on parallel processing*, 2021, pp. 1–10.
[26] B. S. Modjtahedi, J. Wu, T. Q. Luong, N. K. Gandhi, D. S. Fong, and W. Chen, "Severity of diabetic retinopathy and the risk of future cerebrovascular disease, cardiovascular disease, and all-cause mortality," *Ophthalmology*, vol. 128, no. 8, pp.

1169–1179, 2021.
[27] S. Vujosevic *et al.*, "Screening for diabetic retinopathy: new perspectives and challenges," *Lancet Diabetes \& Endocrinol.*, vol. 8, no. 4, pp. 337–347, 2020.
[28] B. E. Kurtul, A. Kurtul, and F. Yalç\in, "Predictive value of the SYNTAX score for diabetic retinopathy in stable coronary artery disease patients with a concomitant type 2 diabetes mellitus," *Diabetes Res. Clin. Pract.*, vol. 177, p. 108875, 2021.