International Journal of Engineering Trends and Technology
ISSN: 2231–5381 / https://doi.org/10.14445/22315381/IJETT-V73I1P124

Volume 73 Issue 1, 288-304, January 2025

Original Article

Comprehensive Meta-Learning Approach for Predicting Gestational Diabetes Mellitus: A Comparative Analysis of Imputation Methods and Classifiers

T. Sujatha[1], K. R. Ananthapadmanaban[2]

[1,2]*Department of Computer Science, SRM Institute of Science and Technology, Vadapalani Campus, Chennai, India.*

[1]*Corresponding Author: tsujathasrm@gmail.com*

Received: 19 September 2024 Revised: 11 January 2025 Accepted: 16 January 2025 Published: 31 January 2025

Abstract *- Gestational Diabetes Mellitus (GDM) constitutes a health risk for everyone during pregnancy; as such, it requires proper early diagnosis and booster. This work offers a detailed analysis of machine learning approaches for GDM prediction, focusing on the interaction between imputation strategies and classification models. The study utilizes a robust benchmark dataset from Kaggle to evaluate the quality and reliability of classifiers, including Decision Stump, Decision Table, Bayes Net, and KNN classifiers, as well as SVM and imputation with both KNN and SVM. We compare the results using performance measures such as accuracy, precision, recall, F1-score, mean cross-entropy, ROC, and PRC scores. The results establish that KNN imputation performs appreciably better than SVM imputation, mainly in terms of prediction accuracy and more even-handed performance for almost all classifiers. Several models integrating KNN imputation with the sophisticated classifier accurately present the performance landscape, with classification accuracy at not less than 97%. Although some SVM-based models' excellent performance shows higher predictive accuracy, they prove to possess greater complexity. This study's results show the importance of choosing the right classifier and imputing missing data when making machine learning models for GDM prediction. By handling the missing data and improving the classification methods, the study provides a direction for a large-scale, accurate, and efficient diagnostic approach, which may significantly improve women's health through prompt and accurate GDM identification.*

Keywords *- Gestational Diabetes Mellitus, Decision Table, KNN Imputation, Decision Stump, SVM Imputation.*

1. Introduction

This disease is now becoming rampant all over the world because millions of expectant women are being diagnosed with GDM each year. The present study indicates that GDM is not only linked to preeclampsia and macrosomia but also predetermines the mother and child for their whole life. The T2DM rate varies from 14% in the developed countries to 9-12% in developing countries. The main reason for the enhancement of gestational diabetes mellitus, as well as many related complications, is early and correct diagnosis of the disease, after which measures can be taken to prevent deterioration of the mother's or baby's health. With data mining coupled with artificial learning algorithms such as Random Forests, Gradient Boosting Machines and even Neural Networks, their utilization in perioperative management methods to predict diseases and the risk level of various diseases has increased, leading to superior process decision-making. Nevertheless, as indicated in Kang et al. (2023), several limitations exist in current studies on the use of ML to predict the risk of GDM; these include the inability to determine a consistent model performance even when applying enhanced algorithms such as the XGBoost. Furthermore, many of the methods tend to center their pattern on stunning architectures that are cumbersome to compile; in most cases, these architectures demand many more resources than can be provided by constrained environments. This study aims to partially fill these gaps by describing a sound meta-learning framework that integrates DI methods. SVM and KNN were chosen due to their complementary strengths: SVM is more suitable for non-linear models, while KNN is ideal when we need to preserve the local structure of a data set, which makes them ideal for clinical datasets- To compare the effectiveness of each method in predicting GDM, SVM, KNN and different other classification methods in ML were applied. The novelty of this research lies in its dual focus: First, to the best of our knowledge, there is no similar work that has scores of (1) assessing the performance of the advanced imputation techniques on the predictive models; second, (2) assessing the trade-off between the complete paradigm of models' accuracy and the time taken to develop those models. All classifiers are compared using the Decision Table, Bayes Net, KNN & SVM, plus SVM & KNN imputation ways. Comprehension of how these methods work in terms of time and space, even when implemented in performance-critical healthcare centres. A combined methodological approach validated on an objective GDM dataset with the indication of the real potential for scaling the approach of imputation methods.

Compared to existing studies, this work provides

- Decision Table, Bayes Net, KNN, and SVM, together with the missing data solution, are SVM and KNN imputations.
- Information about the computational complexity of these methods is important for implementing them in clinical practice.
- A strong methodological foundation, confirmed for a GDM dataset, is available, revealing its scope and repeatability.

This work is organized as follows: Section 2 presents a review of related research, Section 3 describes materials and techniques, Section 4 displays results and analysis, Section 5 concludes the work, and Section 5 provides a description of the future scope.

2. Literature Survey

The previously mentioned related works offer helpful support in the execution of the present study. This section is mainly dedicated to identifying the works related to the present research study. While filling up missing data, Sumathi et al. [1] also used DSAE to delete unwanted events, KNN and HC. The results obtained from this work are as follows: precision: 96.17%, recall: 98.69%, specificity: 89.50%, accuracy: 96.18%, and F-score of 97.41% on the GDM dataset. According to the results, they achieved higher accuracy than decision trees, logistic regression, and neural networks. The process of identifying an outlier wrongly clustered 590 and rightly clustered 2935. Besides, outlawing improves the classification level and allows parents who need a precise diagnosis to be singled out at the initial stage.

The results obtained showed that they outperformed these techniques on the GDM dataset. The model is capable of diagnosing GDM and sub-typing it also. The researchers validated the model through only one self-controlled dataset. Further comparison of the suggested deep learning model to DSAE was not made. Sumathi et al. [2] have stated that class naming, missing-value replacement, and normalization are involved. In categorization, the researchers used random forest, KNN, SVM, and logistic regression. They compared favorably with standard machine learning models, attaining a 94% F-score, 94% precision, 94% recall, and 94.24% accuracy. Applying ensembles makes it possible to make use of several machine learning models. Preparation points to problems with data quality. It helps in the diagnosis of prenatal diabetes at an early enough time. The evaluation may not be as comprehensive as that of deep learning or other methods of the present date.

Khanna K et al. [3] developed an ensemble stack model that includes five approaches of data leveling technique and several configurations of the ML framework. They incorporated an explainable AI layer by integrating the SHAP, LIME, quantum lattice, ELI5, anchor, and feature importance libraries. To enhance the solving of the best gestational diabetes diagnosis model, the researchers trained an ensemble model: SMOTE-ENN. It was 96% specific, 95% for the sensitivity parameter, and 99% for the accuracy determinant. Other predictive factors, including visceral adipose deposits, replaced them. Some advantages include early problem identification, better patient management, and a medical expert’s ability to interpret. The drawbacks of this approach include the issue of bias in the selection of the machine learning models, restrictions in direct patient-doctor interaction, the question of when and how often the models should be updated, the ethical question regarding the use of patients’ data; the extent to which generalisation of the findings can be made depending on the size and variety of the datasets.

Kang B.S. et al. [4] investigates 34,387 pregnancies in seven institutions in South Korea. With XGBoost and LGBM, they made baseline, E0, E1, and M1 GDM prediction models. To derive the predictive features, they used the Boruta algorithm and SHAP values based on all the variables and employed all independent variables, as well as selected simple models. They managed to design easy-to-complete clinical questionnaires that incorporated these variables. There are some problems in the study; they used normal clinical data without biomarkers or genetic factors for classifications, used outdated data, lacked external validation, had missing data, and had possible selection bias and data from only Korean people.

In detail, Cubillos, G. et al. [5] employed 1,611 pregnancies to fine-tune 12 machine learning algorithms to predict GDM. This paper established that data augmentation and selective three-variable selection mechanisms have enhanced the predictive models’ capability. By applying AUCROC, the researchers checked and confirmed model performances against different test sets. When we used four models with seven to twelve input variables, the AUC ROCs reached 0.82, the specificities were 0.72–74 percent, and the sensitivities were 88 percent at most. Other models are characterized by fewer input variables (5) and relatively low specificity (0.62), although the sensitivity of the models amounts to 0.89. This paper has the following limitations: sample size is small; the external validation is missing; there might be some sort of sample selection bias in the dataset; and the method’s performance is compared to the model only, not its interpretability or compared to other methods. Benham, J.L. et al. [6] two comprehensive systematic reviews for gestational diabetes mellitus identified accuracy measures for pharmacotherapy and behavioural interventions that are effective. Analyzing the signal extraction results, we also found several other frequency-specific responses predominantly of precision lifestyle interventions apart from the conventional clinical markers like GDM history, BMI, diagnostic blood glucose levels, etc. Some of the issues are that no precision lifestyle intervention studies have been performed; systematic reviews have been used instead of primary research.

Mennickent, D. et al. [7], employing near-infrared spectroscopy to determine serum samples of the first and second trimesters, utilized four spectral lines and a total of eighty mathematical pre-treatment for each range. Every ML model was cross-validated with double using single and multifactor block validation. The best model was determined to diagnose and use the AUROC of 0.5768 ± 0.0635 during the first trimester and 0.8836 ± 0.0259 in the second trimester. All these results were accomplished in thirty-two minutes and only required 10 µL of serum to generate a GDM prediction.

A retrospective cohort study of 30,474 GDM pregnancies was conducted by Liao, L.D. et al. [8] at Kaiser Permanente Northern California, between 2007-2017. From the discovery cohort EHRs of 2007–2016 and the temporal validation cohort EHRs of 2017, the study found potential predictors at 4 levels (1–4). The model became more understandable when the ensemble (super learner) and the transparent (LASSO regression) machine learning strategies were compared. All time point predictors were used by the super learner to make the best prediction. The limitations of the model include interacting poorly with other HC systems, a poor ability to forecast drug treatments without definitions of the types of treatments, no consideration of how treatment decision biases could affect model conclusions and a lack of information on the actual uses of the super learner model and its potential advantages over other forms of clinical decision.

Kurt B et al. [9] developed the clinical decision support system for 489 patients between 2019 and 2021 using the RNN LSTM and Bayesian optimization. The study aimed to reduce the use of OGTT and to find out the predictors of GD. These include enhanced deeper learning methods, minimization of the side effects due to fewer unnecessary tests, enhanced superior prediction capability, plans for conducting prospective studies with great data, and the creation of clinical decision support. Some issues are associated with this study. It has several limitations, especially because it has a very small number of participants, and it is possible to have low external validity.

Watanabe M et al. employed data obtained from 82,698 expecting mothers of the Japan Environment and Children's Study (JECS) birth cohort [10]. A large sample size increases the statistics and availability rate of data, provides a better comparison for machine learning techniques, discovers the factor of GDM, investigates the decision tree-based algorithm for its correct functioning and meaning and develops the factor of GDM. In their study, Uchitachimoto G. et al.[11] use the method called DC, which resulted in better ROC- AUC (0.767) and recall (0.867) in the checkup data of 324 residents. The study helped me understand some of the GBDT challenges during the processing of secret material. Exploring the accuracy shown by the LR and DC analysis gives almost the same result even with a limited dataset.

Tuppad A. et al. [12] on identifying ML solutions for T2D treatment and prevention. In other words, the study's goal was to provide a prognosis, diagnose non-invasive and invasive features, and assess risk. They searched for literature using the databases PubMed and Google Scholar. The review identified gaps in medical practice, evidence-based guidelines, and knowledge about diabetes. Among the drawbacks of risk scores and models based on laboratory data, it is necessary to include the following factors: the probability of laboratory data is often not directly available and requires additional validation and clarification of its potential and effectiveness in practice.

In pregnancy case-control study by Hu X et al. [13] the testing subjects comprised 190 individuals from August 2020, while the training participants comprised 735 participants from August 2019 to November 2019. In this study, the extreme gradient boosting (XG Boost) machine learning algorithm is based on 20 predictors out of 33 variables. Criteria to evaluate the optimized version of XGBoost were an accuracy of 0.875 and an AUC of 0.946. The AUC and prediction accuracy of a typical logistic regression were 0.752 and 0.786, with four predictors only. In the study, compared with the XGBoost model, the LR model had excellent calibration, but the XGBoost model showed higher discriminative performance.

First, Wu YT [14] analyzed 73 first-trimester characteristics in electronic medical records to predict early GDM onset. Out of the variables identified using feature selection facilitated by machine learning, seven of them are of practical use. For both the full set of 73 variables and the reduced set of 7 variables, we developed models with state-of-the-art machine learning techniques. Surprisingly, total thyroxin T4 and total 3,3,5'triiodothyronine T3 were found to have a higher accuracy as compared to free T3 and T4. Lipoprotein (a) yielded an AUC potential predictive value of 0.66.

Among them, Assi E. et al. [15] have defined the placental proteome of expectant mothers with GDM and LGA newborns. GDM and LGA change the expression of a number of proteins in the placenta, increasing or decreasing the amount of some of them. 1. High expression of DPP-4 and PRG2-bone marrow proteoglycan was observed. 2. DPP-4 and PRG2 changed the markers of stem cell differentiation, so their embryonic function appeared abnormal. New therapeutic interventions for LGA and GDM may be developed from these outcomes. The prospective observational study by Cremona A et al. [16] enrolled pregnant women between 10 and 16 weeks of pregnancy and aged between 18 and 50 years. To control selection bias, we excluded women pregnant with foetuses of less than 18 years of age, twin pregnancies, known foetal abnormalities, and those women with oedema-related issues prior to their pregnancy. Dependent variables were eight-point skinfold thickness (SFT), mid-upper-arm circumference (MUAC), weight, hip and waist circumference, subcutaneous adipose tissue (SAT), and visceral adipose tissue (VAT). BMI, waist, gluteal hip, abdominal SAT, VAT, and truncal SFT were positively associated with gestational diabetes. A multivariate prediction model, comprised of insulin resistance, perinatal mortality and family history of diabetes, could predict GDM (area under curve 0.860). This technique has therapeutic value in GDM risk assessment because it should enable the introduction of effective therapies to ascertain early risks in pregnancy during the first trimester.

Employing the technique of expert interviewing and literature review, the study conducted by Wei LL et al. [17] aimed at pregnant women who are scheduled to undergo antenatal testing for symptoms of GDM. In order to formalize indicators and categorize them, we used the random forest regression method in Python. In light of the study's findings, the study stresses the importance of sample proportions, which are used in the random forest model to indicate GDM and provide an early-stage prognosis of the condition.

In a recent systematic review of studies using machine learning for predicting GDM, Kokori et al. [18] considered published literature between January 2000 and September 2023. To identify cross-sectional themes and patterns, we reviewed 14 papers in detail. Forecasting gestational diabetes mellitus risk in pregnancy, population-specific models and enhanced utilization of clinical data to improve its management.

Ye Y et al. [19] used machine learning and compared it to logistic regression to predict GDM. 3182 women (14.31%) of 22,242 singleton pregnancies had GDM3. These findings identified previous lab results, prior medical conditions, and demographics of the mothers as predictors. They used two regressions and eight frequently employed machine learning algorithms (GBDT, AdaBoost, LGB). We expected GDM5 based on FBG, HbA1c, lipid profile, and BMI, and based on the results, the GBDT model has the best accuracy (AUC 0.74). The risk scores of 0.3 and 0.7 were used to divide the patients into risk classes. The aims of Wang X et al. [20] improved understanding of pregnant women's risk of GDM and the risk factors associated with GDM. This work uses various models to improve the forecast of GDM through ensemble learning. Du, Y. et al. [21]: GDM has a significant impact on the health of expectant mothers and their fetuses since GDM is marked with glucose intolerance. A correlation work of medical and family history to assess the danger of GDM has been provided by Amarnath S et al. [22]. Data-based classification models employ some inference functions based on risk factors and sickness features to predict the relevance of a linked factor. Based on the experimental evidence, it is found that the prognosis model based on the classification may be helpful in the early detection of GDM and can lead to proper management. The advantages include first-chance detection of risks, flexibility in risk profiling, mining and categorization of risks, and integration with emerging technologies such as wearable measures and IoT devices for detecting early signs of illnesses. The absence of first signs, multiple interacting hazard factors, the nature of the data, and its applicability in conditions with scarce resources are challenging.

Mennickent D et al. [23] used several ML approaches for early GDM identification before the standard screening timeframe of 24 – 28 weeks of pregnancy. Machine learning uses various aspects to enhance its efficiency and can potentially predict earlier to allow early action to be taken. These are the following problems: a high proportion of the promising models is not externally validated, they require validation in various populations, the weak predictive ability of the validated models, and difficulties connected with integrating such models in clinical practice – organizational, legal, technical problems.

In the study of Lee SM et al. [24], 1,443 pregnant women who underwent ultrasonography examinations for screening of GDM at 24-28 weeks and NAFLD at 10-14 weeks were selected for a prospective cohort research study. Our models are made with patient basic clinical variables, observations from new guidelines encountered in the recent past, steady variables concerning NAFLD (presence of NAFLD and NAFLD lab findings), and response variables through a stepwise selection method for 11 variables. They used deep neural networks, Support Vector Machine (SVM), Random Forest, and Logistic regression.

Li YX et al. [25] studied two separate cohorts: Of these pregnancies, 2,795 were conducted at Shanghai Pudong New Area People's Hospital (SPNPH) and 4,799 at the Xinhua Hospital Chongming branch (XHCM). They developed prediction models using three machine learning methods, including logistic regression and two ensemble algorithms with severe gradient boosting and 45 first-trimester characteristics. XGBoost model revealed good to excellent performance in the late first trimester (AUC = 0.99), while moderate performance at the initial stage of pregnancy (AUC = 0.75) in the XHCM grouped cohort. In the external validation of the model on the SPNPH cohort, the XGBoost classifier's performance was average (AUC = 0.83). Y. C. Lai and colleagues employed a 75-g OGTT as a GDM one-step screening test method between 24 and 28 weeks of pregnancy.[26] Using first and second trimester clinical risk factors for the mother, such as maternal age and first-trimester BMI, first-degree relative diabetes mellitus and fetal macrosomia, we developed a nomogram and a prediction model by performing multiple logistic regression analyses. The statistical study yielded an AUC of 0.814 (95% CI: 0.-05; 751 - 0.877) using the Windows 10 operating system with Python 3.0 and the SPSS 25. In the model, specificity was 74.5%, and sensitivity was 79.2% when the probability was projected at 0.745.

Zeng C Xie X et al. [27] conducted this prospective cohort study at Zhejiang University School of Medicine'sWomen's Hospital with 721 pregnant women. They also offered screening OGTT during the first trimester of pregnancy for early identification and during 24-28 weeks for the diagnosis of gestational diabetes. Some of the study's limitations include small sample size, single centre study, variations in availability between settings, and the need for longer-term follow-up measures to adequately assess the device's usefulness. In any case, the model may help improve the clinical diagnosis and management of GDM.

In a retrospective cohort study, Guo F et al. [28] examined 3956 women who were first-time doctor seekers in Shanghai in a hospital in Shanghai in 2015 and 6572 in 2016. Biochemical parameters included age, BMI before pregnancy, first-trimester fasting plasma glucose (FPG) and history of diabetes in first-degree relatives. The DCA also sustained its clinical decision-making value beyond the threshold probabilities, confirming a favourable net benefit. Among the advantages we can name are clear visualization, a high degree of clinical relevancy, wide validation, and the possibility of early prediction and decision-making.

From 2016 to 2022, Niu ZR et al. [29], using a retrospective cohort design, there were 4,000 cases in total: 2,975 NGDM and 1,225 GDM. Overall, in the validation cohort of patients, there were 1800 patients – 1281 had NGDM and 519 had GDM. The logistic regression-based prediction model showed a phenomenal performance. The index was highly sensitive and specific, confirming the Receiver Operating Characteristic curve with an Area Under the Curve- AUC value of 0.803 among the modeling cohort, while the AUC for validation was 0.782. First, this model enables the conduct to identify first-trimester indicators that point to high-than-normal-risk pregnancies, as these form the foundation of the overall strategy.

From July 2020 to April 2021, Kang M et al. [30] conducted a prospective cohort study in Shanghai General Hospital to discover the relationship between blood markers and GDM and to establish a nomogram for early pregnancy that will foresee GDM among 413 pregnant women of which 116 were diagnosed with GDM in the follow-up. They quantified age, pre-pregnancy BMI, complication, B lymphocyte percent, FPG, HbA1c, lipids, and progesterone in prim and second-trimester blood samples. We prepared the nomogram using a 5-fold cross-validation and a multivariate logistic model. There are several benefits: statistical analysis is complete, the model is therapeutically used, and the possibility of predicting all the risk factors at the beginning of the process is possible. The study abnormalities include a small sample size, a single center, and a late data collection time, which requires a large and varied set of validations and follow-up research to enhance its usability in clinical practice.

Qingwen et al. [31] The quantity of fatty acid metabolites was measured by assessing GPR120 expression in the white blood cells of the hospital lab; clinical parameters. The present study shows that the level of GPR120 can predict the occurrence of GDM in early pregnancy, thus the need for more participants and long-term monitoring studies.

An automatic third-generation USG with an assessment of first-trimester AST/ALT was used by R et al. [32] in a prospective cohort research study where 666 pregnant women were studied for GDM, of which 94 were positive. The study has certain drawbacks connected with the single-centre approach and the lack of complete validation in the samples that are significantly more numerous and diverse. Therefore, the authors proposed that the AST/ALT ratio in early pregnancy can predict the risk of GDM and recommend more assessments of the AST/ALT ratio's mechanism in predicting GDM and its generalisability.

Liu H et al. [33] employed concrete ML to predict gestational diabetes mellitus (GDM). They studied a logistic model and the XGBoost model. A 7.6% of participants fell under the category of developing GDM, making the XGBoost model outshine the baseline logistic model by having a higher AUC. I also found that the predictions of the test dataset were as expected with the probability calculated by the model.

The first study, by Gao S et al. [34], used a population-based cohort of 19,331 pregnant women from Tianjin, China, to establish a model to predict pre-15 gestational week GDM. Nine potential predictors of early onset diabetes, such as the mother's age, BMI, and history of diabetes among first-degree relatives, were used in the Early Prediction Score based on the first prenatal care visit. Pregnancy-related predictors included in the Comprehensive Score were physical activity, sitting time and passive smoking. In light of agreeing with the study's conclusion of both risk scores as having accurately predicted GDM in pregnant Chinese women, the study calls for further testing in different populations and other strategies.

Wang H et al., in their study [35], reported that high-income countries had higher proportions of GDM at 14.2% than middle-income at 9.2% and low-income countries at 12.7%. Global differences in the distribution of GDM and differences in the prevalence connected with geographical location and income level influence public health initiatives and the allocation of resources.

Many researchers have explored GDM prediction and management using different ML approaches to improve diagnostic ability. Related works, including Random Forest and XGBoost, which are ensemble methods, have reached relatively high predictive accuracy in related fields, while neural networks are slowly but surely being applied to analyze complex data patterns. Below, this section presents an overview of selected works with a particular focus on their methods, conclusions, and weaknesses.

Various techniques applied in handling missing data are still quite a common issue in today's healthcare datasets, and any incorrect or inadequate handling of the missing data may lead to biased or even more unreliable predictions. Sumathi et al. [1,2] used KNN imputation and hierarchical clustering for handling missing values, but KNN was inappropriately for handling high dimensions and showed poor performance for large data sizes. These limitations can be bridged by integrating it with SVM, which is significantly suitable for handling the non-linear distribution of data by ensuring a high pattern probe while at the same time retaining the local structure using KNN, achieving a high precision of 96.17% and recall of 98.69% on GDM data set. The problem with their study is that it did not compare the results of other imputation procedures, and they did not measure computational complexity. Cubillos et al. [6] statistical imputation techniques were applied; however, the authors noted that such methods need to be more powerful in addressing the problem of sparsity when aiming to make predictions based on clinical data. This research extends from these insights by proposing a dual-imputation strategy that combines SVM and KNN to improve prediction accuracy while avoiding datasets' collapse.

ML classifiers for GDM prediction, conventional techniques and ensemble methods Most of the conventional techniques relating to ML classification techniques have been used in the prediction of GDM. In 2024, Varada et al. [3] created an ensemble stack model that included explainability layers of SHAP and LIME to obtain progressive results of 96% precision and 99% accuracy. However, their approach works well when making predictions and is computationally expensive, making it difficult to adopt in settings with few resources. Conversely, Kang et al. [4] used simpler models like XGBoost and LGBM, emphasizing generalizability but reporting lower performance metrics (AUC: 0.804). This work compares the Decision Table and SVM to choose a moderate complexity solution with the balance of Classified accuracy rate coupled with efficiency. The current study also provides a benchmarking comparison of these models with KNN and SVM imputation, which can be used clinically.

Liao et al.'s Clinical Applications and Generalizability Studies [8] and Mennickent et al. 's Clinical Applications and Generalizability Studies [7] concurred in pointing out that there is a need to validate the applicability of templates with sampled population and appropriateness to various research contexts. In order to address these challenges, this research utilized a stratified sampling technique, and the datasets applied in this research can accommodate many demographic populations; therefore, the applicability and accuracy of the introduced predictive models are higher. These two papers, therefore, affirmed the issue of large-scale data and external validation, which are mostly characterized by localized data bias. This research used a publicly accessible GDM database whose characteristics are 3,525 samples and 17 attributes, which are demographic and clinical characteristics like age, BMI, and history of diabetes in first-degree relatives. In a way employed to select training and testing sets that would display such variations, high reliability of the resulting predictive models is attained through stratified sampling. Furthermore, it considers the subsequent confirmation in multisite samples to improve the findings' operational usability.

- Computational cost is another important factor that still contributes to the model being clinically implemented since models with high accuracy tend to consume many resources that come with high costs, thus rendering implementing models in low-resource areas negligible.
- Furthermore, the black-box nature of many presently popular methodologies, like neural networks, negatively affects clinician trust because the models do not reveal how they arrived at particular solutions.
- Insufficient research on imputation methods and their effect on the classifying capability of the classifier.
- Lack of focus on computation complexity required for running models, especially for real-world applications.
- Poor explainability, thus limiting clinician satisfaction and utilization.

3. Materials and Methods

In this research work, a dataset on Gestational Diabetes Mellitus (GDM) obtained from the Kaggle data repository [36] is analyzed. The dataset comprises 3525 instances with 17 attributes, each providing valuable information related to factors potentially associated with GDM.

3.1. Description of the GDM Dataset

To achieve this study's objective, the GDM dataset acquired from the Kaggle data repository was used. It has a total of 3,525 records covered by 17 variables that include basic personal data, diseases, health indicators, and behavior patterns. These include age, BMI, HDL, OGTT, prediabetes and PCOS, physical inactivity, previous pregnancies, gestational complications and birth outcomes. The dependent variable in the current model is categorical, specifically a dichotomy between GDM and non-GDM cases. To minimize the effect of missing data in the subsequent analysis and enhance the modelling techniques' performance, we applied the simple imputation techniques of SVM and KNN. This dataset was shown to have an inherent problem of providing little representation of GDM cases due to class imbalance. To address this problem, data is oversampled so that the size of the minority class is equal to the size of the majority class. Further, an artificial simulation dataset was created by assuming the range of values of the original features, and while creating a new set, their statistical characteristics were maintained. This led to a final cleaning of the data that gave a total of 229 samples, with 115 samples as GDM cases and the remaining 114 samples as non-GDM cases. By using these preprocessing techniques, along with other methods such as stratified sampling and tuning of parameters, we were able to obtain the optimal platform from where to develop highly accurate predictive models of GDM.

3.2. Information Features of the Dataset

3.2.1. Features: 17

- Case Number (Patient Case ID): Numeric
- Age: Numeric, range 20 to 45
- Number of Pregnancies: Numeric, values {0, 1, 2, 3, 4}
- Gestation in previous Pregnancy: Numeric, values {0, 1, 2}
- BMI: Numeric, range 13.3 to 45
- HDL: Numeric, range 15 to 70
- Family History: Numeric, binary {0 = No, 1 = Yes}
- Unexplained prenatal loss: Numeric, binary {0 = No, 1 = Yes}
- Large Child or Birth Default: Numeric, binary {0 = No, 1 = Yes}
- PCOS: Numeric, binary {0 = No, 1 = Yes}
- Systolic BP: Numeric, range 90 to 185
- Diastolic BP: Numeric, range 60 to 124
- OGTT: Numeric, range 80 to 403
- Hemoglobin: Numeric, range 8.8 to 18
- Sedentary Lifestyle: Numeric, binary {0 = No, 1 = Yes}
- Target variable:Outcome(GDM/Non GDM)
- Distribution of the cases: GDM Vs Non-GDM cases

3.3. Description of the Simulation Dataset

This constructed a simulation dataset using SVM and KNN imputation techniques to handle missing values on the features of the dataset in order to improve the resilience of our ML models. This artificial data is used to:

- Expand the sample size by adding to the current data.
- The initial dataset was unbalanced, but then the classes were balanced.
- Verify that our models
- Keeping the original datasets with 17 characteristics intact.
- Producing values for every characteristic that falls within the designated limits.
- To guarantee accurate connections between features, statistical models are utilized.
- In order to attain balance, the minority class (GDM cases) is oversampled.
- There are 229 samples in the final simulated dataset, split 50/50 between GDM and non-GDM situations.

Figure 1 shows that predicting an optimal outcome by using the ML models has the following methods.

3.4. Descriptive Characteristics for Predicting Gestational

3.4.1. Diabetes Mellitus

The factors that are utilized to forecast GDM fall into the following categories

- Demographic: Case number and age
- Pregnancy History: Total Pregnancies, Previous Pregnancy Gestation, Inexplicable Prenatal Death, Large
- Offspring, or Birth Defect
- Health Metrics: Hemoglobin, HDL, OGTT, Systolic and Diastolic Blood Pressure, and BMI
- Medical History: Prediabetes, PCOS, and Family History
- A sedentary lifestyle as a lifestyle factor
- Result: GDM Condition

The paper also discusses the data imputing technique where K-Nearest Neighbors (KNN) and Support Vector Machine (SVM) are mentioned, but the rationale for their consideration is not well explained. There is a need to expand the literature to identify other possible imputing techniques to handle the dataset under consideration. Using basic statistical imputations such as mean median or mode requires less computational work, but the technique may not effectively capture the inherent multivariate interdependence in the GDM dataset due to the mixed type variables present in the data set. Multiple Imputations by Chained Equations (MICE) could be more suitable for more complex issues depending on data type and having complex dependencies.

Likewise, a matrix factorization approach such as SVD is also efficient at practising high-dimensional data, but the data must be balanced and structured. Machine learning can also approach other options, namely the Random Forest Imputation or Deep learning autoencoder techniques. These methods can fit a curve and allow for a wider description of the presence of the relationship between the attributes, which is ideal for forming more complex models to answer the research questions. The authors have limited themselves to stating that KNN and SVM imputation methods match the interest in machine learning models; more explanation as to why the former ones were preferred over the others could have included, for example, the fact that they are more appropriate for continuous and categorical data and outliers as well as since they capitalise on the similarities among the data points. Besides, a brief comparison of their performance against other imputation techniques on the dataset could have supporting empirical arguments for their selection.

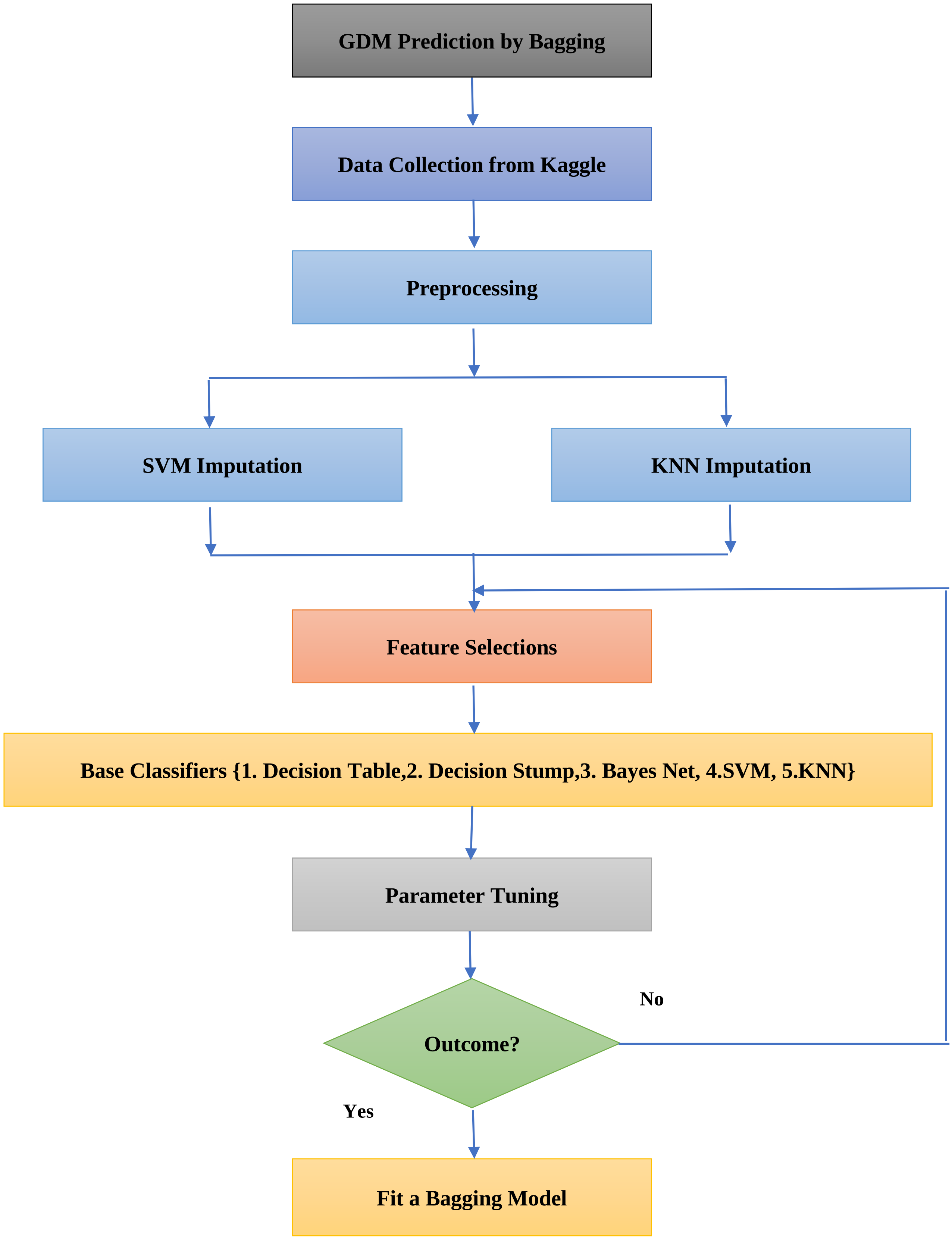

Fig. 1 Schema of the proposed system

Here applied machine learning methods to a thorough risk factor evaluation:

Decision Stump

It is a simple type of decision tree used in supervised learning. It acts as a base classifier in many ensemble methods. It is a one-level decision tree. It makes decisions based on a single feature, creating a simple binary decision rule. Decision stump=a_1,a_2,…a_d, d=no of features and single feature= ai, The decision rule expressed as: a_i≤t.

Decision Table

A decision table is a tabular representation of rules that map input conditions to output actions. It helps us make decisions based on specific combinations of input values. Each row in the table corresponds to a specific combination of input features. The columns represent input features (conditions) and output class labels (action). Let's denote the decision table as DT: DT = {(X1, X2, …, Xn, Y)}.

Bayes Net

A Bayesian network consists of nodes (variables) and edges (dependencies). Let's denote the nodes corresponding to the features as X1, X2,..., Xn. The node representing the class label is Y. Each node has a conditional probability distribution given its parents. the probability of the class label specified all input attributes = P(Y | X1, X2, …, Xn). The probability of attributes Xi given its parents is P(Xi | Parents(Xi)).

K-Nearest Neighbors (KNN)

The technique is a straightforward and efficient method for supervised learning (classification and regression task). It performs on the similarity between K data points. x: The new data point (input features); D: The dataset of existing data points; d(x, xi): The distance between x and each data point xi in D,K: the number of neighbors to consider. Y_hat(x) = mode (Yi) for i in K nearest neighbors (for classification), and Y_hat(x) = mean (Yi) for i in K nearest neighbors (regression).

Support Vector Machine (SVM)

Let's denote the feature vector for a pregnant woman as x= (x1, x2,…xn), Where n= no of relevant features. The SVM aims to find the optimal hyperplane that separates GDM-positive and GDM-negative instances. The hyperplane equation can be expressed as w.x+b=0 and considered a Radial Biased Function Kernel for capturing nonlinear relationships.

Our machine learning models required data to be prepared through a number of important procedures. We used a unique SVM Imputation Technique and KNN imputation to deal with missing values. The dataset was then divided via stratified sampling into an 80% training set and a 20% test set, preserving the original class distribution of GDM and non-GDM cases. By using this method, the training and test sets are guaranteed to accurately reflect the entire dataset. Lastly, we used grid search with ten-fold cross-validation to hyperparameter tune each model to maximize performance. It was possible for us to develop strong and trustworthy machine-learning models for GDM prediction because of this thorough data preparation procedure. The below Pseudocode is considered for this research work to yield a better outcome in predicting GDM for the considered models. Python implements this in colab and Weka 3.8.6 tool for predicting an optimal outcome by using below ML models. The above input features have missing values for BMI, HDL, OGTT, and SysBp features. So, we implemented one of the powerful supervised learning algorithms used for the data imputation methodology. Here, utilised SVM imputer class and KNN imputer by Python- scikit learn to replace missing value.

Pseudocode: DIAMOND: Dual Imputation and Multi-classifier Orchestration for Navigating Diabetes
Input: Gestation Diabetic Data from Kaggle Dataset
Output: Fit a model for predicting Gestation of Diabetic
Step 1: Let X represent the dataset containing n instances (rows) and m features (columns), X= [xij]nxm where xij is the value of feature j in instance i.
Step 2: Let Missing(xij) be a function that returns true if xij is missing
Step 3: M={(i,j)∣Missing(xij)=True}.
Step 4: Let Noise(xij) be a function that returns true if xij is noise
Step 5: N={(i,j)∣Noise(xij)=True}
Step 6: SVM Imputation for X′=X/(M∪N)
Step 7: Identify Step 2, 3, 4, 5, 6 Outcome for X′=X/(M∪N)
Step 8: KNN Imputation for X’′=X/(M∪N)
Step 9: Identify Step 2, 3, 4, 5, and 8 Outcome for X’′=X/(M∪N)
Step 10: Y= Bagging Model ∈ (A, B, C, D, & E) ← X’ & X’’, Where A=SVM, B=KNN, C= Decision Stump, D= Decision Table, and E=Bayes Net
Step 11: Parameter tuning with 90:10 sampling technique
Step 12: Compare (A’, B’, C’, D’, and E’) & (A”, B’’, C’’, D’’, and E’’) Where, A’ & A’’ =outcome of A, B’ & B” = outcome of B, C’ & C’’=outcome of C, D’ & D’’=outcome of D, and E’& E’’=outcome of C

Step 13: Repeat Step 6 to Step 7 until get an optimal solution
Step 14: Fit a Bagging model

Procedure for Data Imputation by SVM and KNN Imputation method
Step1: Start the imputation process
Step 2: Dataset denoted as D: {Y= Y1, Y2…Yn},
Step 3: Compute the Missing dataset denoted as Mij=Yij∈Y, Compute the Complete dataset denoted as Cij=Yij∈Y
Step 4: Target Variable T: Yj, Yj+1, Yj+2…YZ
Step 5: Set Standard scaler
Step 6: Train the model
Step 7: Impute missing values to T
Step 8: Integrate from step 7
Step 9: repeat step7 and 8 until Mij=0
Step 10: Stop

This work governs the classification evaluation metrics and regression evaluation metrics below.

$$\text{Accuracy} = \frac{(\text{TP}+\text{TN})}{(\text{TP}+\text{TN}+\text{FP}+\text{FN})} \quad (1)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP}+\text{FP}} \quad (2)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP}+\text{FN}} \quad (3)$$

$$\text{False Positive Rate} = \frac{\text{FP}}{\text{FP}+\text{FN}} \quad (4)$$

$$\text{F1} - \text{Score} = \frac{2*(\text{Precission}*\text{Recall})}{(\text{Precision}+\text{Recall})} \quad (5)$$

$$\text{MCC} = \frac{(\text{TP}*\text{TN}-\text{FP}*\text{FN})}{\sqrt{(\text{TP}+\text{FP})*(\text{TP}+\text{FN})*(\text{TN}+\text{FP})*(\text{TN}+\text{FN})}} \quad (6)$$

$$\text{Kappa Statistic} = \frac{2*(\text{TP}*\text{TN}-\text{FP}*\text{FN})}{(\text{TP}+\text{FP})*(\text{TP}+\text{FN})*(\text{TN}+\text{FP})*(\text{TN}+\text{FN})} \quad (7)$$

ROC curve= It is plotted with the TPR on the y-axis and the FPR on the x-axis

PR curve= It is plotted with PPV values on the y-axis and TPR values on the x-axis

$$\text{Mean Absolute Error} = \frac{1}{\text{n}}\sum_{\text{i}=1}^{\text{n}}|\text{x}_\text{i} - \text{m(X)}| \quad (8)$$

Here, m(X)=average value of the data, n=no of data, and xi=data values

$$\text{Root Mean Square Error} = \frac{\sqrt{\sum_{\text{i}=1}^{\text{Num}}\|\text{y(i)}-\hat{\text{y}}\text{(i)}\|^2}}{1^{\text{Num}}} \quad (9)$$

Here, Num= No of data points, $y(i) = i$th measurement, and $\hat{y}(i)$ =corresponding prediction.

$$\text{Relative Absolute Error} = E_i = \frac{\sum_{j=1}^{n}|P_{ij}-T_j|}{\sum_{j=1}^{n}|T_j-\bar{T}|} \quad (10)$$

$$\text{Root Relative Square Error} = E_i = \frac{\sum_{j=1}^{n}|P_{ij}-T_j|2}{\sum_{j=1}^{n}|T_j-\bar{T}|2} \quad (11)$$

Here, TP=True Positive, TN=True Negative, FN=False Negative, FP=False Positive.

4. Results and Discussion

In this section, the authors present the result and evaluation of Gestational Diabetes Mellitus (GDM Data Set). Therefore, the bagging model was used with the selected base learning algorithms, namely, SVM, KNN, Bayes Net, Decision Table, and Decision Stump, using the SVM imputation and KNN imputation. The accuracy scores are presented in Table 1 for the different combinations of imputation methods (SVM and KNN) with different classifiers.th selected base learning algorithms the SVM, KNN, Bayes Net, Decision Table and Decision Stump through SVM imputation and KNN imputation on this Gestational Diabetes Mellitus (GDM Data Set) to identify the best model.

Table 1 presents accuracy scores for various combinations of imputation methods (SVM and KNN) with different classifiers. KNN Imputation has slightly outperformed SVM Imputation in most classifiers tested. The lowest accuracy of the Decision Stump classifier for both imputation techniques is estimated at 87.43%. While it is observed that some of the classifier combinations yield accuracy lower than 94%, other combinations yield accuracies that are higher than 94%. The three combination models are KNN Imputation + KNN classifier, KNN Imputation + SVM and KNN Imputation + Bayes Net with an accuracy of 97.19%, 97.10% and 96.96% respectively. It is seen that SVM Imputation gives better results when integrated with Bayes Net (95.83%) and Decision Table (95.52%).

Table 1 shows the Precision values of the study using combinations of imputation methods (SVM and KNN) with classifiers. The Decision Stump classifier has the lowest performance with the lowest Precision of 0.88 for SVM

Imputation and 0.89 for KNN Imputation. Comparing the results of both imputation techniques with different classifiers used, KNN Imputation performs better or equal to SVM Imputation in every case. The highest Precision score of 0.97 is found between KNN Imputation and Decision Table, Bayes Net, KNN and SVM classifiers. When SVM Imputation has been included with Decision Table and Bayes Net, two combinations give the best result where the Precision is 0.96. Both the SVM Imputation + KNN and the SVM Imputation + SVM give nearly the same Precision score of 0.95. The difference in performance with different methods of imputing is least when the Decision Stump classifier is used (0.01), and it is maximum when KNN and SVM classifiers are used (0.02). In general, it is observed that KNN Imputation gives a higher or at least equal Precision score than SVM Imputation, where the difference ranges from .01 to .02 points.

Table 1 shows Recall values obtained for combining the different imputation techniques with SVM and KNN with the different classifiers. The Recall is identified to be the lowest for the Decision Stump classifier for both SVM and KNN Imputation. The KNN Imputation, together with the classifiers of Decision Table, Bayes Net, KNN and SVM, yields the best Recall score of 0.97. SVM Imputation when used in conjunction with decision table and Bayes Net, two methods produce a recall of 0.96. For the combination of SVM Imputation + KNN and SVM Imputation + SVM results, a Recall of 0.95 was attained. As seen in Table 7, the disparity in performance between imputation methods is largest for the KNN and SVM classifiers (0.02). Altogether, for all the datasets, the Recall of KNN Imputation is higher or, at least, equal to the Recall of SVM Imputation with an increase of 0.01 to 0.02 points. This further supports the fact that using KNN-based imputation is favorable for this specific dataset or problem. The Recall values align with the Precision values discussed earlier; this again suggests a fair degree of accuracy concerning the true positive rate and the positive predictive value. These outcomes agree with the formerly reported performance, steadily indicating that KNN Imputation outperforms other methods according to different criteria.

Table 1 shows the ROC (Receiver Operating Characteristic) scores – an informative measure of the performance of different classifications derived from imputations and classifiers. The ROC curve area measurements are between 0 and 1, where completely separated classes have higher AUC elad. All models in this comparison have high ROC values of 0.99, indicating near-perfect classifiers. This top performance is observed for all combinations using KNN imputation apart from KNN imputation with Decision Stump. SVM imputation also marks high, with ROC figures of 0.991 with Decision Table and Bayes Net classifiers. The ROC score of 0.84 is the lowest for both SVM and KNN imputation where the Decision Stump – classifier was used, showing that the latter performed poorly with all the imputed sets. The average accuracy results with KNN and SVM classifiers show that KNN imputation slightly outperforms the SVM imputation, where the accuracy was 0.99/0.98/0.97 for KNN classifier/SVM classifier/KNN imputation. In sum, KNN imputation is found to marginally outperform or have similar performance to SVM imputation across all classifiers examined, while the choice of classifier greatly influenced the imputation model, with Decision Stump being the worst among classifiers tested.

Table 1 shows the PRC scores, giving yet another view of how different model combinations with SVM and KNN imputation methods combined with different classifiers perform. PRC scores are identified as ROC scores that vary in between the range of 0 to 1, where the higher the score, the better the performance. The PRC measures for most of the proposed models in this comparison look perfect, with PRC scores of either 0.98 or 0.99, which denote high precision and recall levels at various probability thresholds. The best overall results with PRC scores of 0.99 comprise both SVM and KNN imputation when using the Decision Table and Bayes Net classifiers. Specifically, after the KNN imputing and adding KNN and SVM classifiers, we have the results of 0.98 and 0.97 for SVM imputing, KNN, and SVM classifiers, respectively. The PRC score of 0.83 is obtained when both SVM and KNN were imputated by the Decision Stump classifier, similar to the ROC score PRC score proves that the end of the alley Decision Stump is the worst classifier irrespective of the imputation technique. In totality, the scores of the PRC are slightly higher than that of the ROC scores as was observed earlier that KNN imputation is slightly better or equally good than SVM imputation in most classifiers. The results also reveal that the type of classifier affects a model most remarkably, whereby Decision Stump appears to be the least effective, while Decision Table and Bayes Net are the most effective classifiers in this study using all imputation approaches, but especially when employing all imputations.

Table 1 shows the Kappa score of the given combinations of the imputations, such as SVM and KNN, with the different classifiers. The Decision Stump classifier has the lowest value of Kappa 0.72 for both the SVM and KNN Imputation methods. By comparing the overall accuracy results for KNN Imputation and SVM Imputation. It can also be seen that the overall Max Kappa scores of 0.94 belong to KNN Imputation with Bayes Net and KNN classifiers. SVM Imputation appeared to do well jointly with the Decision Table and had the Kappa of 0.91. KNN Imputation also has fairly good performance with other classifiers; Decision Table and SVM both have an accuracy of 0.93. The SVM Imputation + KNN combination has the lowest Kappa of all the non-Decision Stump models). Altogether, there are slight improvements of either the same or even higher Kappa scores resulting from KNN Imputation as compared to those from SVM Imputation; thus, it may be inferred that KNN imputation might be the better approach to impute this specific dataset or to solve this particular problem in terms of inter-observer agreement.

Table 1 summarizes the F-Measure values obtained for the current experiments based on the use of the SVM and KNN imputation techniques with a number of classifiers. The F-Measure is derived from the element of precision and recall; it allows equal weight to both measures of a model. The Decision Stump classifier got the lowest F-Measure (0.87 for both SVM and KNN Imputation), meaning that it was the worst of all the options. KNN Imputation performs favorably or better than SVM Imputation in all classifiers. The best F-Measure of 0.97 is realized from KNN Imputation if the recognition algorithms employed by the Decision Table, Bayes Net, KNN and Super Vector Machine are used. The imputation technique SVM achieves the best result of F-Measure of 0.96 when combined with the imputation technique Bayes Net. The two models, SVM Imputation + KNN and SVM Imputation + SVM give an F-Measure of 0.95.

The largest difference in performance between the Imputation Methods is visible in the Decision Table classifier, in which KNN Imputation (0.97) has a 0.03-point advantage over SVM Imputation (0.94). In general, cross validation average F-Measure for KNN Imputation is higher or similar to the one for SVM Imputation, with KNN Imputation obtaining 0.01-0.03 higher F-Measure. As a result, further support is provided to KNN's imputation efficacy for this particular type of dataset or problem. Based on the results of the F-Measure, these scores correlate with the previous lack and recall scores to offer a comprehensive view of the models and evidence that, once again, KNN Imputation delivers the highest performance rates according to all parameters. In Table 1, the values have been reported to compare the performances of different models based on cross-validation imputations combined with classifiers. The top-performing model obtained an MCC of 0.94 and includes KNN Imputation together with Bayes Net, KNN classifier and SVM. At the same time, the non-significant model is the SVM Imputation with MCC 0.74, whereas the second model also has MCC 0.74, but it includes KNN Imputation and Decision Stump.

Interestingly enough, KNN imputation always performs significantly better or similar to the SVM imputation regardless of the chosen classifier. In the classifiers, Decision Stump gave relatively poor results compared to other classifiers; however, Bayes Net, KNN and SVM classifiers, especially if used with KNN imputation, yielded high results. The Decision Table classifier performs well, particularly when paired with KNN imputation (MCC: 0.93). Based on the results described above, it can be concluded that the overall performance of the KNN imputation method is qualitatively higher, and the choice of the classifier influences the quality of the result more than others, with higher classes providing higher results compared to simpler classes such as Decision Stump.

Table 1 below shows how MAE achieved different imputation methods (SVM and KNN) using different classifiers. Poor results are evident when using the Decision Stump classifier with the highest MAE value of 0.22 for SVM Imputation and 0.21 for KNN Imputation. Thus, it is seen that in terms of recall, the KNN Imputation performs better on par with other SVM Imputations, depending on the classifier chosen. The best minimum average error attains 0.02 with the combination of KNN Imputation and KNN classifier, seconded by KNN Imputation combined with Bayes Net and SVM (each attaining 0.03).

The best results are obtained with the combination of Bayes Net and SVM Imputation and SVM Imputation, with MAE = 0.05 each. Here, we also see that for the Decision Table classifier, both imputation methods yield an almost indiscernible 0.09 in MAE. In total, it can be seen that KNN Imputation generates lesser or at least comparable MAE scores with SVM Imputation, which again indicates the efficiency of KNN Imputation for this particular dataset or problem. The relative MAE scores correspond to previous accuracy and Kappa results; therefore, the superior performance of KNN-based imputation methods is confirmed.

Table 1. Classification and Regression Metrics

S.No	Classifier	Accuracy	Precision	Recall	ROC	PRC	Kappa	F1-Score	MCC	MAE	RMSE	RAE	RRSE	Time
1	SVM Imputation+ Decision Stump	87.43%	0.88	0.87	0.84	0.83	0.72	0.87	0.74	0.22	0.33	45.38%	67.28%	0.03
2	KNN Imputation+ Decision Stump	87.43%	0.89	0.87	0.84	0.83	0.72	0.87	0.74	0.21	0.32	45.14%	67.27%	0.06
3	SVM Imputation+ Decision Table	95.52%	0.96	0.96	0.99	0.99	0.91	0.94	0.91	0.09	0.18	21.01%	38.91%	10.08
4	KNN Imputation+ Decision Table	96.88%	0.97	0.97	0.99	0.99	0.93	0.97	0.93	0.09	0.15	18.60%	32.72%	5.92
5	SVM Imputation+ Bayes Net	95.83%	0.96	0.96	0.99	0.99	0.91	0.96	0.91	0.05	0.19	11.24%	39.33%	0.59
6	KNN Imputation+ Bayes Net	96.96%	0.97	0.97	0.99	0.99	0.94	0.97	0.94	0.03	0.17	6.31%	33.98%	0.06
7	SVM Imputation+KNN	94.98%	0.95	0.95	0.98	0.97	0.89	0.95	0.89	0.06	0.2	12.67%	40.06%	0.03
8	KNN Imputation+KNN	97.19%	0.97	0.97	0.99	0.98	0.94	0.97	0.94	0.02	0.15	6.07%	30.47%	0.01
9	SVM Imputation+SVM	95.43%	0.95	0.95	0.97	0.97	0.9	0.95	0.9	0.05	0.19	10.42%	39.61%	228.53
10	KNN Imputation+SVM	97.10%	0.97	0.97	0.99	0.98	0.93	0.97	0.94	0.03	0.15	6.43%	31.69%	106.53

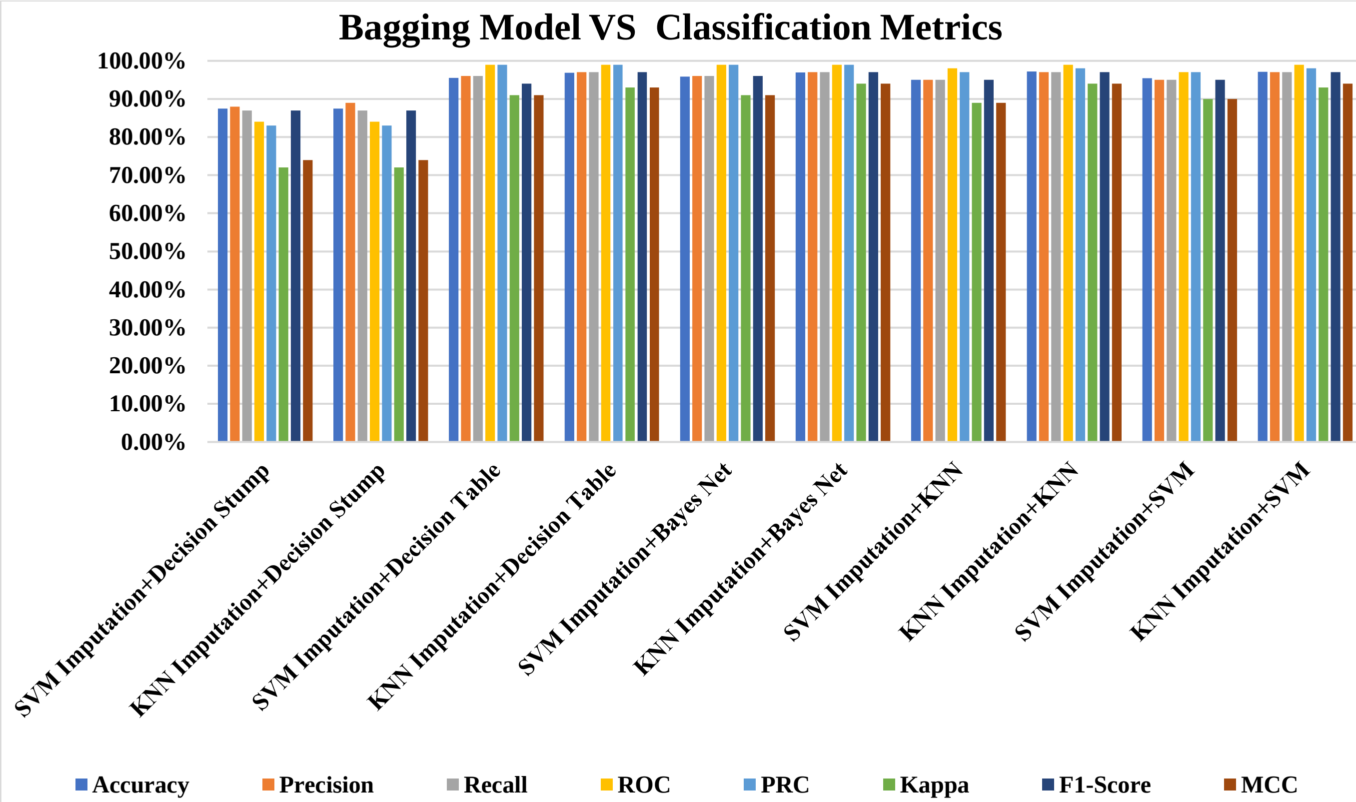

Fig. 2 Graphical representation of Models Vs Classification metrics

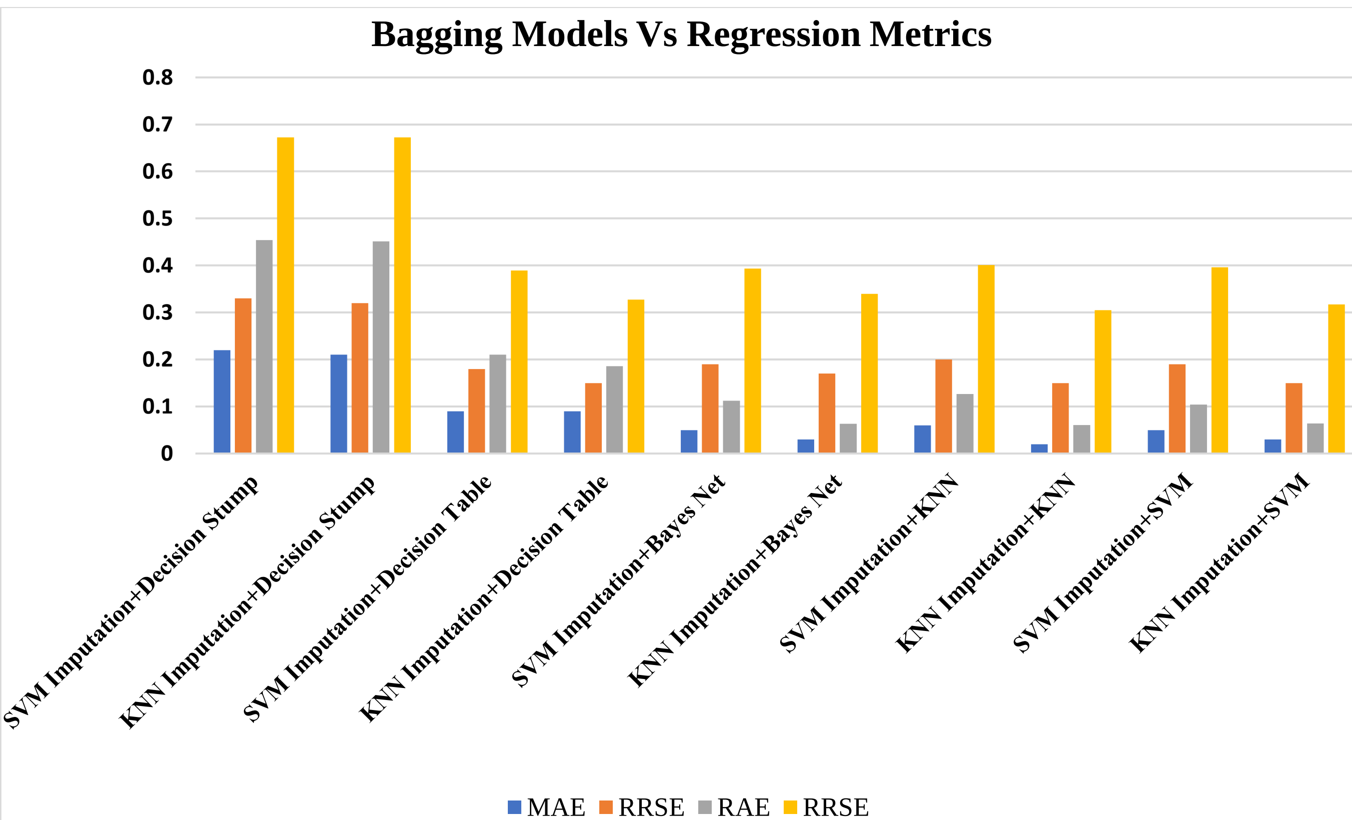

Fig. 3 Graphical representation of Models Vs Regression metrics

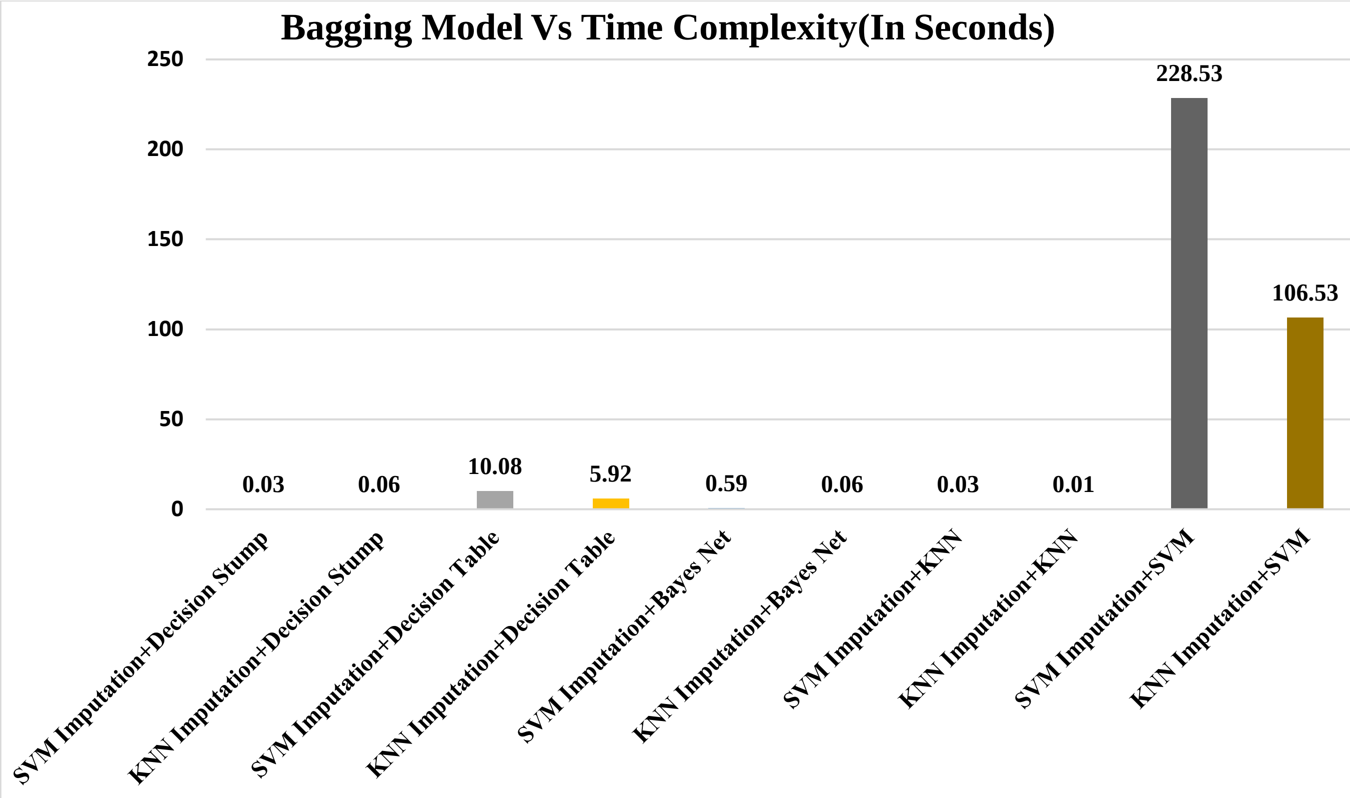

Fig. 4 Graphical Representation of Models Vs Time Complexity

Table 1 shows Root Mean Square Error scores resulting from combinations of imputation methods (SVM and KNN) with different classifiers. The worst performance, as revealed by the highest RMSE values of 0.33 and 0.32, was achieved using the Decision Stump classifier for SVM Imputation and KNN Imputation, respectively, in the classification matrix. KNN Imputation performs better in all classifiers than the other methods, including SVM Imputation. The smallest RMSE equal to 15 percent is predicted for the KNN Imputation accompanied by Decision Table, KNN, and SVM. SVM Imputation yields the highest improvement with Decision Table (RMSE of 0.18) next is Bayes Net-SVM (RMSE of 0.19). The difference between the dialects/RMs of the two methods is biggest with the KNN classifier at 0.20 for SVM Imputed vs 0.15 for KNN Imputed. The RMSE of KNN Imputation is generally significantly smaller than that of SVM Imputation. This confirms the significant advantages of using the specified method for the further particular dataset or problem. These scores of RMSE are consistent with previous means (accuracy, Kappa, MAE) and have shown that KNN-based imputation methods outperform other forms of data imputation across all measures of error totals.

Table 1 provides the Relative Absolute Error percentage derived from the results of applying imputation methods (SVM and KNN) with different classifiers. The last one is the Decision Stump classifier, where we have the lowest RAE (45.38% for SVM Imputation and 45.14% for KNN Imputation). The overall results provide KNN Imputation wins over SVM Imputation for each of the classifiers experimented with. The KNN Imputation + KNN classifier recorded the lowest RAE at 6.07%, followed by KNN Imputation + Bayes Net at 6.31% and KNN Imputation + SVM at 6.43%. SVM Imputing proves to be most effective when paired with SVM (10.42 % RAE), thereafter with Bayes Net (11.24% RAE) and KNN (12.67% RAE). For the Decision Table classifier, there is an improvement in KNN Imputation (18.60% RAE) compared to the SVM Imputation (21.01% RAE). In general, RAE KNN Imputation achieved better results than SVM Imputation for all the classifiers, while the difference in the result between the two imputation methods is somewhat larger for higher-performing classifiers.

Table 1 shows root relative squared error percentages when using either the SVM or KNN imputation method with a range of classifiers. The lowest values are represented by The Decision Stump classifier with 67.28% of RRSE for SVM Imputation and 67.27% for KNN Imputation, which means the worst performance in all configurations. In general, KNN Imputation performed the best and was better than SVM Imputation in all classifiers used. This shows that imputing by KNN combined with KNN classifier gives the best result with a minimum RRSE of %. 3047 while imputing by KNN and SVM and Decision Table follow the next best result with %.3169 and %.3272RRSE, respectively. SVM Imputation gives a modest predictive performance in the case of other non-decision Stump classifiers with RRSE of between 38.91% and 40.06%, respectively, as shown below; the Largest difference between imputation methods is obtained with the KNN classifier, where the RRSE achieved with the SVM Imputation is 40.06 percent compared with only 30.47 percent achieved with the KNN Imputation. In general, KNN Imputation appears to get lower RRSE scores than SVM Imputation across all the classifiers, by between 5-10% for the superior classifiers. This further supports the efficacy of KNN-based imputation for this given dataset or the considered problem. The obtained RRSE scores concord with previous error metrics that indicate the effectiveness of KNN Imputation across different error measures. The time complexity information generated reveals and measures the efficiency of different combinations of SVM and KNN imputation methods in tandem with several classifiers. These times expressed in seconds are the exact time taken to train each of these models. The quickest to train is KNN Imputation+KNN with 0.01 secs and also is outdone by SVM Imputation+KNN, which only took 0.03 secs, as well as SVM Imputation+Decision Stump, which took 0.03 secs. The KNN Imputation, together with Decision Stump and Bayes Net, also has a relatively short training time of 0.06 seconds each. However, it takes the longest time to train the model that uses SVM as a classifier out of all the models addressed in this research; the model takes 228.53sec to train SVM Imputation+SVM while, at the same time, KNN Imputation+SVM took 106.53sec. This implies that SVM classifiers, though quite efficient in performance indicators, are accompanied by considerable computation burden. Training time in the Decision Table classifier is also relatively high, namely, 10.08 for the SVM Input + Decision Table and 5.92 seconds for the KNN Imput+Decision Table. However, SVM Imputation+Bayes Net cost half an hour (0.59 sec) at least five times longer than the cost of KNN Imputation+Bayes Net, 0.06 sec.

In general, it is evident from this data that while some models, like the ones using the SVM classifiers, are highly accurate, most of them are very slow to train compared to the models using the KNN or the Decision Stump classifiers.

5. Conclusion

This work concludes that the current systematic review of the effectiveness of applying the ML algorithm in diagnosing GDM has yielded significant outcomes of significant therapeutic value. The study repeatedly demonstrated that the K-Nearest Neighbors (KNN) imputation is more advantageously placed compared to other techniques in terms of classifiers and performs ant measures, achieving amazing accuracy levels of up to 97.19 %. Specifically, it was observed that techniques like Bayes Net, KNN, and SVM, which were more complex classifiers, generated higher accuracy than the other straightforward classifiers when combined with KNN imputation. Nevertheless, this study also highlighted the fact that higher accuracy attached to the proposed family of classifiers, such as SVM-based classifiers, comes at a heavy price of very long training times compared to their less accurate counterparts.

These models seem to perform well consistently across different evaluation measures, and these results support GDM risk assessment as a useful tool with the possibility of early intervention that leads to better maternal and fetal health. For future research, the relative importance of these individual predictive features should be investigated, and other aspects for increasing the model accuracy have to be examined. These models have to be tested in clinical sample groups. This work can be considered a valuable contribution to the development of goal-oriented and more effective methods in preventing obstetric complications, with the possibility of a step change in the approach and management of GDM by identifying targets for early intervention.

References

[1] A. Sumathi, S. Meganathan, and B. Vijila Ravisankar, “An Intelligent Gestational Diabetes Diagnosis Model Using Deep Stacked Autoencoder,” *Computers, Materials & Continua*, vol. 69, no. 3, pp. 3109-3126, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[2] A. Sumathi, and S. Meganathan, “Ensemble Classifier Technique to Predict Gestational Diabetes Mellitus (GDM),” *Computer Systems Science and Engineering*, vol. 40, no. 1, pp. 313-325, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[3] Varada Vivek Khanna et al., “Explainable Artificial Intelligence-Driven Gestational Diabetes Mellitus Prediction Using Clinical and Laboratory Markers,” *Cogent Engineering*, vol. 11, no. 1, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[4] Byung Soo Kang et al., “Prediction of Gestational Diabetes Mellitus in Asian Women Using Machine Learning Algorithms,” *Scientific Reports*, vol. 13, no. 1, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[5] Gabriel Cubillos et al., “Development of Machine Learning Models to Predict Gestational Diabetes Risk in the First Half of Pregnancy,” *BMC Pregnancy Childbirth*, vol. 23, no. 1, pp. 1-18, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[6] Jamie L. Benham et al., “Precision Gestational Diabetes Treatment: A Systematic Review and Meta-Analyses,” *Communications Medicine*, vol. 3, no. 1, pp. 1-13, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[7] Daniela Mennickent et al., “Simple and Fast Prediction of Gestational Diabetes Mellitus Based on Machine Learning and Near-Infrared Spectra of Serum: A Proof of Concept Study at Different Stages of Pregnancy,” *Biomedicines*, vol. 12, no. 6, pp. 1-20, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[8] Lauren D. Liao et al., “Development and Validation of Prediction Models for Gestational Diabetes Treatment Modality Using Supervised Machine Learning: A Population-Based Cohort Study,” *BMC Medicine*, vol. 20, pp. 1-13, 2022. [CrossRef] [Google Scholar] [Publisher

Link]

[9] Burçin Kurt et al., "Prediction of Gestational Diabetes Using Deep Learning and Bayesian Optimization and Traditional Machine Learning Techniques," *Medical & Biological Engineering & Computing*, vol. 61, pp. 1649-1660, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[10] Masahiro Watanabe et al., "Prediction of Gestational Diabetes Mellitus Using Machine Learning from Birth Cohort Data of the Japan Environment and Children's Study," *Scientific Reports*, vol. 13, no. 1, pp. 1-11, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[11] Go Uchitachimoto et al., "Data Collaboration Analysis in Predicting Diabetes from A Small Amount of Health Checkup Data," *Scientific Reports*, vol. 13, no. 1, pp. 1-8, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[12] Ashwini Tuppad, and Shantala Devi Patil, "Machine Learning for Diabetes Clinical Decision Support: A Review," *Advances in Computational Intelligence*, vol. 2, pp. 1-24, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[13] Xiaoqi Hu et al., "Prediction Model for Gestational Diabetes Mellitus Using the XG Boost Machine Learning Algorithm," *Frontiers in Endocrinology*, vol. 14, pp. 1-10, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[14] Yan-Ting Wu et al., "Early Prediction of Gestational Diabetes Mellitus in the Chinese Population via Advanced Machine Learning," *The Journal of Clinical Endocrinology & Metabolism*, vol. 106, no. 3, pp. e1191-e1205, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[15] Emma Assi et al., "Placental Proteome Abnormalities in Women with Gestational Diabetes and Large-for-Gestational-Age Newborns," *BMJ Open Diabetes Research & Care*, vol. 8, no. 2, pp. 1-8, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[16] Alexandra Cremona et al., "A Risk-Prediction Model Using Parameters of Maternal Body Composition to Identify Gestational Diabetes Mellitus in Early Pregnancy," *Clinical Nutrition ESPEN*, vol. 45, pp. 312-321, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[17] Li-Li Wei et al., "Application of Machine Learning Algorithm for Predicting Gestational Diabetes Mellitus in Early Pregnancy," *Frontiers of Nursing*, vol. 8, no. 3, pp. 209-221, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[18] Emmanuel Kokori et al., "The Role of Machine Learning Algorithms in Detection of Gestational Diabetes; A Narrative Review of Current Evidence," *Clinical Diabetes and Endocrinology*, vol. 10, pp. 1-15, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[19] Yunzhen Ye et al., "Comparison of Machine Learning Methods and Conventional Logistic Regressions for Predicting Gestational Diabetes Using Routine Clinical Data: A Retrospective Cohort Study," *Journal of Diabetes Research*, vol. 2020, pp. 1-10, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[20] Wang Xiaojia et al., "Analysis and Prediction of Gestational Diabetes Mellitus by the Ensemble Learning Method," *International Journal of Computational Intelligence Systems*, vol. 15, pp. 1-20, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[21] Yuhan Du et al., "An Explainable Machine Learning-Based Clinical Decision Support System for Prediction of Gestational Diabetes Mellitus," *Scientific Reports*, vol. 12, pp. 1-14, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[22] Sumathi Amarnath, Meganathan Selvamani, and Vijayakumar Varadarajan, "Prognosis Model for Gestational Diabetes Using Machine Learning Techniques," *Sensors and Materials*, vol. 33, no. 9, pp. 3011-3025, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[23] Daniela Mennickent et al., "Machine Learning-Based Models for Gestational Diabetes Mellitus Prediction before 24-28 Weeks of Pregnancy: A Review," *Artificial Intelligence in* Medicine, vol. 132, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[24] Seung Mi Lee et al., "Nonalcoholic Fatty Liver Disease and Early Prediction of Gestational Diabetes Mellitus Using Machine Learning Methods," *Clin Mol Hepatol*, vol. 28, no. 1, pp. 105-116, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[25] Yi-xin Li et al., "Prediction of Gestational Diabetes Mellitus at the First Trimester: Machine-Learning Algorithms," *Archives of Gynecology and Obstetrics*, vol. 309, pp. 2557-2566, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[26] Bruno Basil et al., "A First Trimester Prediction Model and Nomogram for Gestational Diabetes Mellitus based on Maternal Clinical Risk Factors in a Resource-Poor Setting," *BMC Pregnancy and Childbirth*, vol. 24, pp. 1-8, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[27] Shan Wu et al., "A Prediction Model of Gestational Diabetes Mellitus Based on OGTT in Early Pregnancy: A Prospective Cohort Study," *The Journal of Clinical Endocrinology & Metabolism*, vol. 108, no. 8, pp. 1998-2006, 2003. [CrossRef] [Google Scholar] [Publisher Link]

[28] Fei Guo et al., "Nomogram for Prediction of Gestational Diabetes Mellitus in Urban, Chinese, Pregnant Women," *BMC Pregnancy and Childbirth*, vol. 20, no. 1, pp. 1-8, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[29] Z.R. Niu, L.W. Bai, and Q. Lu, "Establishment of Gestational Diabetes Risk Prediction Model and Clinical Verification," *Journal of Endocrinological Investigation*, vol. 47, no. 5, pp. 1281-1287, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[30] Mei Kang et al., "A Novel Nomogram for Predicting Gestational Diabetes Mellitus During Early Pregnancy," *Frontiers in Endocrinology*, vol. 12, pp. 1-8, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[31] Qingwen He et al., "Predictive Value of First-Trimester GPR120 Levels in Gestational Diabetes Mellitus," *Frontiers in Endocrinology*, vol. 14, pp. 1-9, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[32] Rongjing An et al., "AST-to-ALT Ratio in the First Trimester and the Risk of Gestational Diabetes Mellitus, *Frontiers in Endocrinology*, vol. 13, pp. 1-9, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[33] Hongwei Liu et al., "Machine Learning Risk Score for Prediction of Gestational Diabetes in Early Pregnancy in Tianjin," *Diabetes/Metabolism Research and Reviews*, vol. 37, no. 5, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[34] Si Gao et al., "Development and Validation of an Early Pregnancy Risk Score for the Prediction of Gestational Diabetes Mellitus in Chinese Pregnant Women," *BMJ Open Diabetes Research & Care*, vol. 8, no. 1, pp. 1-11, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[35] Hui Wang et al., "IDF Diabetes Atlas: Estimation of Global and Regional Gestational Diabetes Mellitus Prevalence for 2021 by International Association of Diabetes in Pregnancy Study Group's Criteria," *Diabetes Research and Clinical Practice*, vol. 183, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[36] Gestational Diabetes Mellitus (GDM) Data Set, Kaggle, 2021. [Online]. Available: https://www.kaggle.com/datasets/sumathisanthosh/gestational-diabetes-mellitus-gdm-data-set