

Prifysgol Cymru
Y Drindod Dewi Sant
University of Wales
Trinity Saint David

Module Title: Research Methods and Professional Development.

Module Code: ACFA7004

Assignment Title: Research Proposal.

Student Name: MD Mahadi Hasan

Student ID: 2505591

Programme:

MSc in Software Engineering and Artificial Intelligence

University Name: University of Wales Trinity Saint David (Swansea Campus)

Lecturer:

Seena Joseph

Submission Date: 8th May 2026

Table of Contents

1	Introduction	3
1.1	Context and Background	3
1.2	Research Problem.....	4
1.3	Research Aim	4
1.4	Research Objectives	4
1.5	Significance of the Research	5
2	Literature Review	6
2.1	Introduction to the Literature Review	6
2.2	Machine Learning Approaches to GDM Prediction	7
2.3	Deep Learning Versus Classical Machine Learning - The Interpretability Argument	8
2.4	Explainability Frameworks and Clinical Decision Support.....	8
2.5	Feature Selection and Dataset Challenges	9
2.6	Literature Summary.....	10
3	Research Design and Methodology.....	10
3.1	Research Design.....	10
3.2	Research Method and Strategy.....	12
3.2.1	System Architecture and Design.....	13
3.2.2	Methods of Data Analysis and Evaluation.....	14
3.3	Research Materials	14
3.3.1	Research Data	14
3.3.2	Ethical Considerations	15
3.3.3	Software, Libraries and Hardware.	15
4	Anticipated Results and constraints.	16
5	Project Timeline.....	16
6	References.....	18

1. Introduction

1.1 Context and Background

Gestational diabetes mellitus (GDM) is a type of glucose intolerance that develops or is initially recognized during pregnancy and is one of the most clinically important metabolic complications of the contemporary obstetric environment. Over the past years, machine learning (ML) has opened up as a paradigm of transformational analytical tools in clinical informatics, which can be used to identify complex, non-linear relationships in large datasets of patients, which are ill-equipped to capture by the traditional statistical tools [1]. The breadth and momentum of the ML based GDM research is reflected in the results of a systematic meta-analysis of twenty-five published ML prediction studies, which reported a pooled area under the receiver operating characteristic curve (AUROC) of 0.8492 across a wide variety of algorithms and datasets - a figure that substantially supports the high predictive potential of the field [2]. The clinical stakes are accordingly high: GDM women are seven to ten times more likely to develop Type 2 Diabetes Mellitus (T2DM) in their later life [3], making it an early GDM prediction priority that extends far beyond the gestational period itself.

The standard clinical test in the diagnosis of GDM is the Oral Glucose Tolerance Test (OGTT) which may be administered between 24 and 28 gestation weeks. Although the OGTT is still the diagnostic gold standard, its timing creates a structural and preventable gap: the physiological markers of risk of developing GDM could be measured earlier in pregnancy but the traditional diagnostic window is not opened until the second trimester. Cubillos et al. demonstrated that the early prediction of the variables is feasible and clinically significant with the optimized ML models achieving sensitivity of 0.82 and an AUROC of 0.81 at the beginning of pregnancy [4]. These findings confirm the existence of a clinically significant proportion of GDM cases, in principle, predictable earlier than current standard practice would permit and that the predictive information exists within routinely collected antenatal data. New trends have swept the field with growing use of ensemble classifiers, gradient boosting architectures and explainability models such as SHapley Additive exPlanations (SHAP) which break down model outputs into single-feature contributions - a broader directional shift in clinical AI towards transparent and clinician-facing decision support.

1.2 Research Problem

With the above progress notwithstanding, the GDM prediction literature can still be characterized by a continued presence of methodological limitations which prevent the scientific and clinical applicability of the literature. Mehrnoush et al. applied seven ML classifiers to a retrospective clinical cohort and found that even high-capacity boosting models only achieved moderate AUC values in the range of 0.64 to 0.66 when applied to real-world datasets comprising noisy, incomplete and highly categorical variables [5], demonstrating the difference between controlled benchmark performance and the messier reality of clinical data. In a scoping review of seventeen articles which included forty-four prediction models, Kirkwood et al. derived the conclusion that the great majority of models lacked external validation and readiness to implement in clinical situations and further identified the absence of multi-class classifiers which could distinguish between pharmacological treatment requirements as a key unresolved gap [6]. Rossi et al. also found additional structural weaknesses in the broader diabetes prediction community such as an over-reliance on the Pima Indians Diabetes (PIMA) dataset, excessive use of binary, rather than calibrated, probabilistic outputs and a systematic failure to address class imbalance all of which reduce the clinical utility of published models [7]. Taken collectively, these restrictions constitute a clear, actionable, and interpretable ML prediction artefact of early GDM that applies principled class-balancing, systematic multi-algorithm comparison and an integrated SHAP explainability to make predictions meaningful and actionable to clinical practitioners.

1.3 Research Aim

The aim of this research is to develop, train, and comparatively evaluate a machine learning-based prediction model for early gestational diabetes mellitus using clinical and demographic features, with the specific objective of identifying the most accurate and interpretable classifier suitable for early clinical risk stratification.

1.4 Research Objectives

The following four SMART objectives operationalize the research aim:

Objective 1 (O1): To critically analyze existing machine learning and deep learning approaches to gestational diabetes mellitus prediction, identifying methodological trends, comparative performance benchmarks and research gaps through a systematic review of published evidence.

Objective 2 (O2): To design and develop an ML-based prediction artefact comprising Logistic Regression, Random Forest and XGBoost classifiers, trained on preprocessed and SMOTE-balanced clinical data derived from the PIMA Indians Diabetes Dataset.

Objective 3 (O3): To evaluate the developed prediction models using standard performance metrics - Accuracy, F1-score, AUC-ROC, Precision, Recall - and to apply SHAP analysis to interpret the contribution of individual clinical features to each classifier's predictions.

Objective 4 (O4): To compare the evaluated model results and identify the optimal GDM prediction classifier with explicit assessment of the trade-off between predictive accuracy and clinical interpretability to determine the most suitable artefact for potential clinical deployment.

1.5 Significance of the Research

The importance of this study is multidimensional at the clinical, technological and public health levels. Clinically, early diagnosis of GDM is well-documented to be risky to both maternal and neonatal outcomes and evidence has shown that early intervention is initiated earlier in pregnancy and yields measurably better health outcomes. A mixed-methods investigation of the adherence to postpartum screening among women previously diagnosed with GDM found that only 40% of women followed up according to the clinically recommended six-week follow up after diagnosis [8], demonstrating that the gap between diagnosis and sustained clinical engagement that the previously, ML-facilitated risk stratification could help to bridge. In a Delphi-based European consensus study, Bajramagic et al. found that AI-based clinical decision support systems have a significant potential to enhance the diagnosis of diabetes, maximize patient personalized monitoring and decrease the inertia of therapy in routine care pathways [9]. In addition to gestational implications, there are emerging genetic findings that have linked GDM with a higher risk of developing ER-negative breast cancer in the long term through Mendelian Randomization analysis [22], which further highlights the longer-term clinical significance of early diagnosis and treatment. This study will create a useful decision-support instrument with results that are easy to understand and apply in real clinical settings by non-expert healthcare professionals by generating a validated, interpretable ML artefact with the results elucidated by SHAP feature interpretations.

2. Literature Review

2.1 Introduction to the Literature Review

The review is a synthesis of the published evidence on machine learning methods to predict gestational diabetes mellitus and includes the period of 2022 to 2026. These four thematic areas include machine learning approaches and their comparative performance, deep learning architecture versus classical ML, explainability frameworks and clinical decision support and feature selection strategies and dataset limitations. The sources were found in IEEE Xplore, PubMed, Scopus, and BMC Open Access databases by using the search terms that combined 'gestational diabetes prediction' and 'machine learning GDM' and 'SHAP explainability diabetes' and 'clinical decision support diabetes'. The review is thematically structured to develop a cumulative argument to support the methodological decisions taken in this study. Thematic overview Table 1 offers a systematic thematic overview.

Theme	Key Sources	Key Findings	Gap Addressed
ML-based GDM Prediction	[10], [11], [12], [13]	RF and XGBoost achieve AUC 0.80-0.94; SHAP explainability is emerging as a methodological standard for clinical deployment	Class-imbalance; single-center datasets; limited external validation
Deep Learning vs Classical ML	[14], [15]	DL achieves 96-98% accuracy but suffers from interpretability deficits and requires dataset volumes exceeding available GDM data	Interpretability deficit; dataset scale constraints; unsuitable for n<1000 GDM datasets
Explainability & CDSS	[7], [16]	SHAP produces actionable feature explanations; probability calibration required for meaningful clinical risk communication	Uncalibrated binary outputs; black-box predictions unsuitable for clinical practitioners

Feature Selection & Datasets	[17], [18], [19]	Feature selection improves accuracy; PIMA dataset has documented demographic restrictions and zero-value imputation requirements	PIMA demographic restrictions; generalizability constrained to Pima Indian female population aged over 21
---	------------------	--	---

Table 1: Thematic Summary of Literature Review

2.2 Machine Learning Approaches to GDM Prediction

The most directly pertinent area of literature relates to the use of classical ML classifiers to early prediction of GDM using first-trimester and early-pregnancy clinical data. The ensemble-based stacking model created by Zaky et al. is trained on first-trimester data of 138 pregnant women and it reports an approximate 89% accuracy with explainability to clinicians via SHAP-based outputs [10]. Although these findings are positive, the overall generalizability of the study is limited by its single-center Egyptian cohort and relatively small sample size - a limitation that is shared by many GDM prediction studies. The cross-sectional study by Bigdeli et al. involved significantly more records of 16,730 patients across six ML algorithms, showing that Random Forest had the best predictive performance of 89% accuracy and AUC of 0.94 [11]. The authors described this superiority to the ensemble averaging mechanism of the Random Forest which reduces the risk of overfitting as compared to single-tree classifiers - a finding which is also associated with the more general literature on ensemble learning. In the contextually specific setting of low-resource antenatal care in Uganda, Ahishakiye et al. applied the Random Forest and XGBoost algorithms to 3,525 clinical records and achieved ROC-AUC values of 0.997 with SHAP analysis showing BMI, HDL cholesterol, blood pressure and pre-diabetes status as the most influential predictors [12]. Although the near-perfect values of AUC raise some questions when it comes to overfitting, what is commendable about this study is its explicit prioritization of explainability as a requirement of clinical deployment - a principle that is shared by this study.

Another issue that comes out amongst this literature strand is the aspect of integrating prediction models within the clinical decision support systems. Using data collected during a randomized controlled trial Du et al. were able to develop an explainable ML-based clinical decision support system with AUC-ROC values reaching 0.792 [13]. The research has

explicitly defined the rejection of black-box predictions as a clinical necessity and made SHAP not an optional analytical accoutrement but a requirement to be responsible about deployment. Within this theme, there is convergent and very strong evidence that both Random Forest and XGBoost are the most highly performing classical ML classifiers to predict GDM and that SHAP-based explainability is rapidly becoming a methodological standard.

2.3 Deep Learning Versus Classical Machine Learning - The Interpretability Argument

The methodological issue that the study will be critically interested in is whether deep learning (DL) architectures are a better choice as opposed to classical ML to predict GDM. A comparative analysis of five DL models: CNN, RNN, LSTM, FNN, and GRU on the PIMA dataset reported that RNN had the highest accuracy of 96.8% and simultaneously identified challenges of class imbalance, model interpretability and lack of external clinical validation that remained unsolved [14]. Al Reshan et al. have created a stack-ANN ensemble, which combines ANN, CNN, and LSTM layers, with accuracy scores of over 98 percent on general diabetes data; however, it is limited with respect to its applicability to the prediction of GDM specifically because of its reliance on non-GDM data and by the complexity of its architecture which makes SHAP-based feature explanation substantially more computationally intensive than with classical ML.

The paper has chosen to take a classical method of ML and the reasons supporting this choice are threefold. To start with, the size of the available dataset (PIMA, n=768) is too small to train deep learning architectures to the consistent convergence they exhibit with severe overfitting. Second, the SHAP is intrinsically compatible with classical ML classifiers (especially Random Forest and XGBoost) to provide clinical-grade feature importance outputs without any post-hoc approximation layers. Third, the literature on GDM prediction uniformly shows that classical ML can be used with AUC values of 0.80 to 0.94 under properly controlled conditions [10, 11, 12], which is performance of a standard that does not require the data demands and interpretability trade-offs of deep learning to achieve.

2.4 Explainability Frameworks and Clinical Decision Support

Explainability has become a distinguishing and defining approach methodological criterion in the clinical AI literature. Rossi et al. created the D.R.E.A.M. framework - a combination of Random Forest, XGBoost and LightGBM with integrated SHAP and probability calibration - achieving AUC values of over 0.83 and F1-scores of over 0.78, with Random Forest offering

the best balance between recall (0.89) and interpretability [7]. The paper directly criticized the wider prediction literature as lacking calibration and interpretability of binary predictions that can then be used to communicate the uncertainty of a prediction and the clinical reasoning as to why a particular model is used to make a specific prediction, as required to make responsible clinical decisions. Using a LightGBM regression framework with SHAP analysis, Ahmed et al. applied continuous glucose monitoring data to a LightGBM regression framework and demonstrated how feature-level explanations (identifying pre-meal glucose slope, baseline glucose and circadian variability as dominant predictors) could be operationalized into individualized clinical guidance [16]. The unanimity in this theme is that SHAP converts opaque outputs into structured, actionable explanations that have meaning to frontline healthcare practitioners - the principle that directly grounds the evaluation design of this research.

2.5 Feature Selection and Dataset Challenges

The most impactful but often overlooked aspects of disease prediction with the help of ML include feature selection and dataset quality. Chan et al. showed that non-traditional biomarkers namely elemental concentrations of nickel, copper and selenium in fingernails could achieve an ensemble AUROC of 0.81 to predict GDM in a nested case-control study [17], demonstrating the potential of feature expansion beyond traditional clinical biomarkers. The sample size of sixty-seven participants was the main limitation of the study as it demonstrates the ongoing conflict between novelty of biomarkers and their statistical power. Sirmayanti et al. systematically tested metaheuristic feature-selection algorithm combinations with ML classifiers on three diabetes datasets and found combinations of feature-selection algorithms with ML classifiers to consistently yield high prediction accuracy by removing irrelevant variables and reducing computational burdens, with Firefly Algorithm combined with Naïve Bayes reaching 92.88% accuracy in the Vanderbilt dataset [18]. El-Bashbishy et al. explicitly addressed the limitation of datasets used by constructing a composite Diabetes Types Dataset that combines four heterogeneous sources including PIMA that achieved 99.98% accuracy on a multiclass classification task and treated the PIMA dataset as restricted to adult women over twenty-one and also required a lot of zero-value imputation [19]. These issues are directly faced in the data preprocessing pipeline of the current study.

2.6 Literature Summary

In the four themes discussed, some of the common patterns are identified. The most trustworthy classical ML algorithms in the GDM prediction are Random Forest and XGBoost which under fairly controlled conditions can achieve AUC values of 0.80 or higher. SHAP explainability is becoming a methodological precondition to clinical use and no longer an analytical option. Deep learning architectures achieve greater raw accuracy but do not have the interpretability properties and data-volume requirements to be realistic at early GDM prediction. The imbalance in the dataset and excessive dependence on the PIMA dataset are all issues that are yet to be addressed by the field in all its structural issues. The current study is also placed to fill these gaps by integrating comparative ML evaluation with SMOTE class-balancing and SHAP explainability into a reproducible Design Science Research framework both to provide a validated prediction artefact and a methodologically transparent prototype of clinical GDM risk stratification.

3. Research Design and Methodology.

3.1 Research Design

The current study is a positivist research philosophy, which is compatible with the epistemological position that the risk of GDM is an objective, measurable phenomenon that is controlled by quantifiable clinical variables that exist irrespective of researcher interpretation. Within positivism, the generation of the knowledge is achieved by means of the empirical observation, systematic measurement, as well as the reproducible testing of the predictive hypotheses by means of the structured numerical data - conditions that are fully satisfied by the experimental design of the current study [21]. A deductive research methodology is used: the study will start with theoretically proposed propositions that are established in the literature on GDM prediction in particular: that glucose concentration, body mass index, maternal age, and insulin are some of the most predictive clinical characteristics - and tests these propositions through controlled computational experimental research, bringing established theory to specific empirical conclusions.

The overall research approach is Design Science Research (DSR), a well-established framework in information systems and computer science research and is characterized by the design and development of purposive artefacts - in this case the ML prediction model - as the primary mechanism by which knowledge is generated and problems solved.

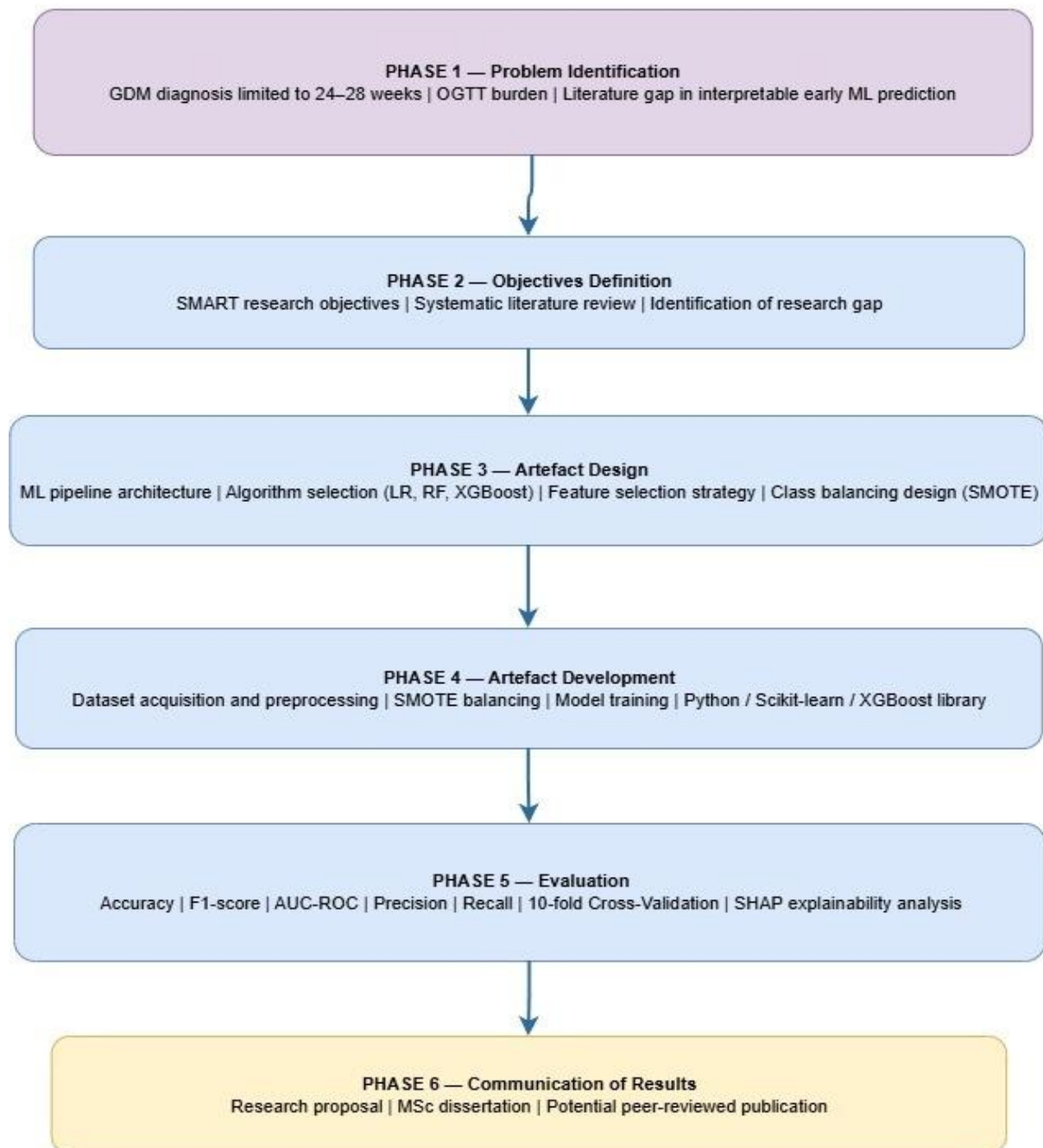


Figure 1 DSR Methodology Flowchart

The cycle used in this research is shown in Figure 1 (DSR Methodology Flowchart), and consists of six consecutive stages: problem identification, objectives definition, artefact design, artefact development, evaluation, and presentation of findings. This model is suitable since the research problem is explicitly solution-focused: it is inadequate to state that the GDM

prediction is a phenomenon that needs to be replicated using the research-produced artefact to the clinically meaningful and reproducible standard.

3.2 Research Method and Strategy.

This research is based on quantitative mono-method research. All data are quantitative; analysis is done by statistical calculation and evaluation of performance metrics and hypothesis of research is tested by objective and reproducible experimental methods. The research strategy is experimental: three ML algorithms are run on the same preprocessed data under the same conditions of cross-validation and their results are systematically compared to determine the best classifier to predict early GDM. This experimental design reflects the methodology of Bigdeli et al., who cross-validated a PIMA-based protocol to provide a directly comparable methodological benchmark [11]. Islam et al., whose cross-validated PIMA-based protocol provides an exactly comparable methodological benchmark [20].

The three classifiers chosen: Logistic Regression (LR), Random Forest (RF), and XGBoost are a purposeful and supported methodological development. LR offers an easy-to-use, linear, transparent baseline with direct interpretability of coefficients, frequently used as a reference standard in clinical prediction literature. RF brings in ensemble averaging of the multiple decision trees, reducing the risk of overfitting and modelling non-linear interactions between features that LR would not be able to model. XGBoost implements sequential gradient boosting - repeatedly refining the error of the prediction of the previous trees - and has shown repeatedly excellent performance in GDM prediction across a variety of independent studies [12, 7]. This three-classifier comparative design allows results to be directly compared to published results. The conscious decision of classical ML, as opposed to deep learning, is fully supported in Section 3.3.

3.2.1 System Architecture and Design

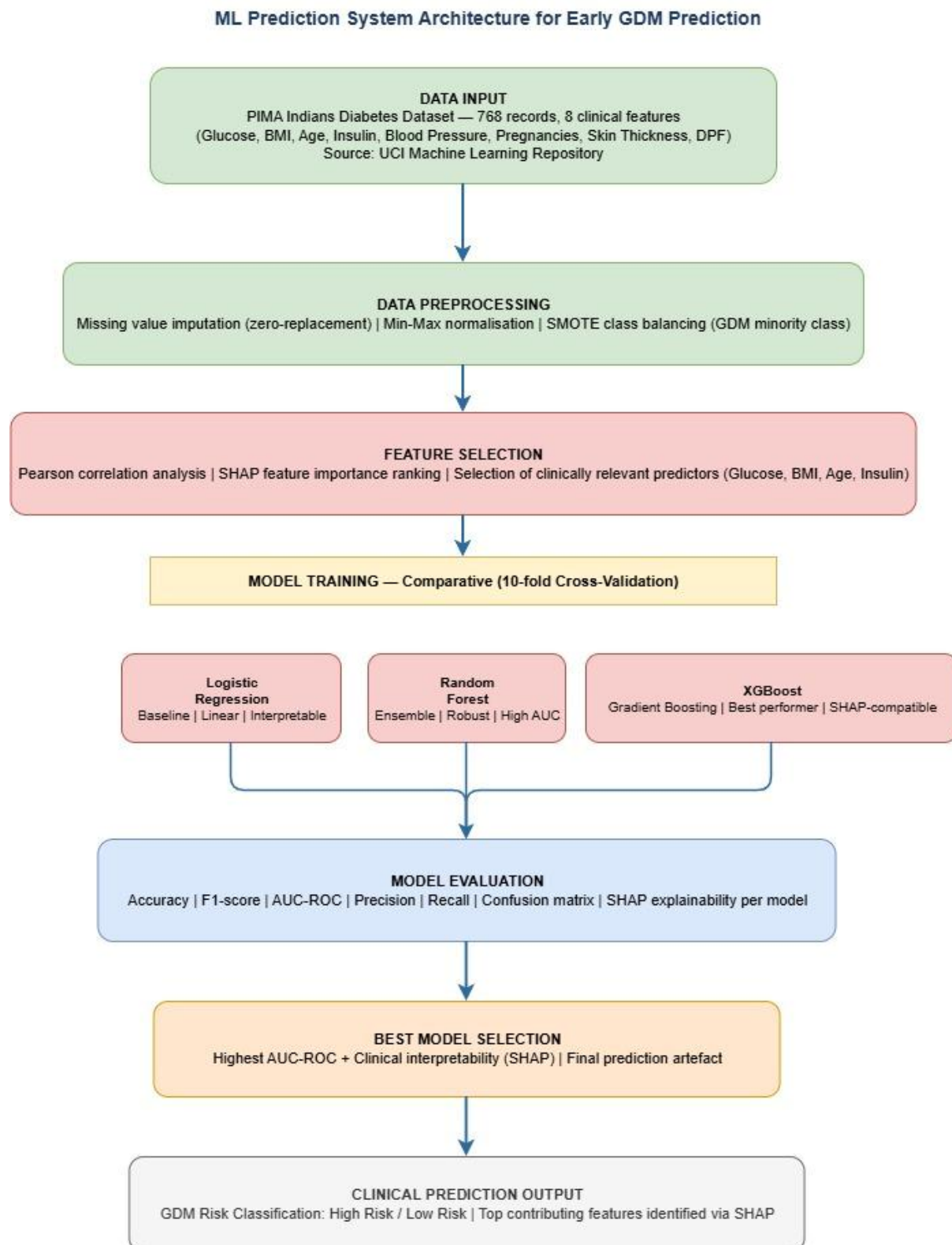


Figure 2 ML System Architecture diagram

The architecture of the ML prediction system, as depicted in Figure 2 (ML System Architecture diagram) is made up of seven sequential steps. It starts with the ingestion of the PIMA Indians Diabetes Dataset, then proceeds to data preprocessing, specifically, median imputation of the physiologically implausible zero values in glucose, insulin, BMI, and blood pressure variables, Min-Max normalization to the [0-1] range and SMOTE-based oversampling of the minor GDM category to address distributional imbalance - and then applies SHAP-guided feature selection to identify the clinical predictors that have the highest level of discriminative value. The three classifiers are trained using ten-fold stratified cross-validation under tenfold parallel training, where the performance is measured against five performance metrics and an optimal model is selected based on the highest AUC-ROC along with the most interpretable clinically SHAP profile. The ultimate result is a GDM risk classification - High Risk or Low Risk - with a ranked SHAP feature importance plot showing which clinical variables motivated the individual prediction.

3.2.2 Methods of Data Analysis and Evaluation.

Five standard measures will be used to measure model performance and are consistently based on the GDM prediction literature: Accuracy, F1-score, AUC-ROC, Precision, and Recall. The first evaluation criterion is AUC-ROC, which gives a threshold-independent quantification of discriminative power and is resistant to class imbalance and is directly comparable across existing studies [10, 11]. The F1-score is chosen as the second criterion, since it balances Precision and Recall a balance that is of critical clinical importance in GDM prediction, where false negatives (undetected GDM cases) have significantly more serious clinical consequences than false positives. All three classifiers are subjected to ten-fold stratified cross-validation so as to obtain performance estimates that will be stable and not artefacts of a particular train-test partition. Each model will calculate SHAP values to produce both a global ranking of feature importance to a model and a single explanation of prediction by the model.

3.3 Research Materials

3.3.1 Research Data

The main data source used in this research is the PIMA Indians Diabetes Dataset that is found in the UCI Machine Learning Repository and that has a total of 768 patient records and eight clinical variables: number of pregnancies, plasma glucose concentration, diastolic blood

pressure, triceps skinfold thickness, serum insulin, BMI, diabetes pedigree function (DPF), and age. The dataset consists of 268 positive GDM cases (34.9%), and 500 negative cases (65.1%), which is a moderate level of class imbalances that will be mitigated using SMOTE in the preprocessing stage [20]. This dataset is the most thoroughly validated and freely available clinical diabetes benchmark that is currently available and can be used directly in similar published prediction studies [7, 11, 19], hence allowing results of this study to be meaningfully compared to that benchmark. Even though the limitation of the dataset to Pima Indian women over twenty-one and the use of physiologically unrealistic zero values in some features are recognized weaknesses [19], they are methodically overcome by using median imputation and normalization as discussed in Section 4.3.

3.3.2 Ethical Considerations

Since the data used in this study will be a publicly available, fully anonymized dataset obtained via the UCI Machine Learning Repository, the researcher will not have to recruit any human subjects and the risks involved will be low. No personally identifiable information is present in the dataset, and all the data processing will be based on the principles of GDPR. The form of ethics is also submitted autonomously in line with module requirements.

3.3.3 Software, Libraries and Hardware.

Python 3.x is the main programming language that is used in the implementation environment. Scikit-learn (implementation of Logistic Regression, Random Forest, and cross-validation pipeline), XGBoost (implementation of gradient boosting) and SHAP (explainability analysis and feature importance visualization) are the core ML libraries. Pandas and NumPy are used to manipulate and preprocess the data and Imbalanced-learn is used to provide the SMOTE implementation of oversampling. The Matplotlib and Seaborn are used as visualization. The computing platform is Google Collab - a cloud-based, free-tier, GPU-enabled platform - which removes hardware acquisition needs and full experimental reproducibility through shareable, version-controlled Python notebooks. Classical ML does not need any dedicated server or cloud computing spend to run at this scale of data.

4. Anticipated Results and constraints.

It is hoped that this study will reveal that either Random Forest or XGBoost will come out as the best prediction model of choice with an AUC-ROC of at least 0.85 on the PIMA data under ten-fold cross-validation. SHAP analysis is predicted to determine glucose concentration, BMI, and maternal age as the top 3 most clinically significant predictors, in line with what is reported in the reviewed literature [12]. The practical implication of these expected outcomes is a validated, reproducible prediction artefact which may, with suitable clinical validation and regulatory review, serve as an early triage instrument in the standard antenatal screening workflows-facilitating targeted OGTT allocation, earlier dietary/pharmacological intervention and less specialist obstetric workload. A clear, benchmarked comparative analysis directly addressing the gaps in class-imbalance and interpretability gaps identified in the literature would also be beneficial to the healthcare informatics community.

There are three main limitations that limit the scope of this study. The limitation to Pima Indian women over twenty-one years old limits the ethnic and geographic generalizability of any model trained on it; findings cannot be assumed to directly transfer to other populations unless external clinical validation is done on independent cohorts. Second, classical ML classifiers, although interpretable and computationally efficient, might not be as effective as deep learning architectures in capturing the full mechanistic complexity of the biological an etiology of GDM. Third, synthetic minority samples as generated by SMOTE may not represent accurately the actual clinical distribution of early-stage GDM cases and might provide inflated estimates of performance which need external validation to substantiate, and which should be interpreted with due caution.

5. Project Timeline

The research project will be done over fourteen weeks starting with literature consolidation and ending with final submission. The timeline is put together to enable an iterative refinement at key methodological stages - especially between model training and evaluation - and to provide enough time to enable revision of write-up and supervisor feedback. Table 2 shows the entire project timeline with relevant milestones.

Week	Phase	Key Tasks	Milestone
1-2	Literature Review	Systematic database search (IEEE, Scopus, PubMed, BMC); source annotation; gap analysis using literature matrix.	Literature database finalized.
3	Research Design	DSR framework finalization; ethics form completion and submission; PIMA dataset acquisition and initial inspection.	Ethics form submitted; dataset secured.
4-5	Data Preprocessing	Median imputation for zero values; Min-Max normalization; SMOTE class balancing; exploratory data analysis (EDA).	Preprocessed, balanced dataset ready.
6	Feature Selection	Pearson correlation analysis; preliminary SHAP importance ranking; final feature set confirmed and locked.	Feature set confirmed.
7-8	Model Development	LR, RF, and XGBoost training; 10-fold stratified cross-validation pipeline implementation in Scikit-learn and XGBoost.	Three trained classifier models.
9	Model Evaluation	Accuracy, F1-score, AUC-ROC, Precision, Recall computation; confusion matrices generated; SHAP value plots produced per model.	Evaluation metrics recorded.
10	Comparative Analysis	Cross-model performance comparison against published benchmarks; optimal model selection; interpretability vs accuracy trade-off assessed.	Best prediction model identified.
11-12	Proposal Write-Up	Full draft of all proposal sections; Figure 1 and Figure 2 integration; complete	Complete proposal drafts ready.

		IEEE referencing audit across all 22 confirmed sources.	
13	Review and Revision	Supervisor feedback integration; Turnitin similarity check; academic language proofreading; final formatting and section word count verification.	Revised, submission-ready document.
14	Final Submission	Final page formatting and style check; ethics form verification via Moodle; Turnitin submission of proposal report by deadline.	Submitted.

Table 2: 14-Week Research Project Timeline

6. References

- [1] E. 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, no. 18, 2024, doi: 10.1186/s40842-024-00176-7.
- [2] Z. Zhang et al., “Machine Learning Prediction Models for Gestational Diabetes Mellitus: Meta-analysis,” *Journal of Medical Internet Research*, vol. 24, no. 3, e26634, 2022, doi: 10.2196/26634.
- [3] A. Prashanthan and J. Prashanthan, “Interpretable Deep Learning for Type 2 Diabetes Risk Prediction in Women Following Gestational Diabetes,” *Scientific Journal of Engineering Research*, vol. 2, no. 1, 2026, doi: 10.64539/sjer.v2i1.2026.376.
- [4] G. Cubillos et al., “Development of machine learning models to predict gestational diabetes risk in the first half of pregnancy,” *BMC Pregnancy and Childbirth*, vol. 23, no. 469, 2023, doi: 10.1186/s12884-023-05766-4.
- [5] V. Mehrnoush et al., “The power of machine learning models in predicting gestational diabetes mellitus,” *BMC Pregnancy and Childbirth*, vol. 26, 2026, doi: 10.1186/s12884-026-08856-1.

- [6] J. R. Kirkwood, N. Galloway, R. S. Lindsay, A. Manataki, D. J. Wake, and R. M. Reynolds, "The use of machine learning to predict pharmacological therapy in gestational diabetes: A scoping review," *Diabetic Medicine*, vol. 43, e70171, 2026, doi: 10.1111/dme.70171.
- [7] D. Rossi, A. A. Citarella, F. De Marco, L. Di Biasi, H. Zheng, and G. Tortora, "D.R.E.A.M: diabetes risk via explainable AI modeling," *Multimedia Tools and Applications*, vol. 85, 2026, doi: 10.1007/s11042-026-21240-7.
- [8] Y. Jyotsna et al., "A Mixed-Method Study on Perception Toward Postpartum Follow-Up Screening for Diabetes among Women Diagnosed with Gestational Diabetes Mellitus in Warangal District, Telangana," *Journal of Pharmacy and Bioallied Sciences*, vol. 18, Suppl. 1, pp. S123-S125, 2026, doi: 10.4103/jpbs.jpbs_946_25.
- [9] M. Bajramagic et al., "Artificial intelligence-driven clinical decision support systems to assist healthcare professionals and people with diabetes in Europe at the point of care: a Delphi-based consensus roadmap," *Diabetologia*, vol. 69, pp. 259-273, 2025, doi: 10.1007/s00125-025-06601-5.
- [10] H. Zaky et al., "Machine learning based model for the early detection of Gestational Diabetes Mellitus," *BMC Medical Informatics and Decision Making*, vol. 25, no. 130, 2025, doi: 10.1186/s12911-025-02947-3.
- [11] S. K. Bigdeli et al., "Predicting Gestational Diabetes Mellitus in the first trimester using machine learning algorithms," *BMC Medical Informatics and Decision Making*, vol. 25, no. 3, 2025, doi: 10.1186/s12911-024-02799-3.
- [12] E. Ahishakiye, J. Nakirijja, and S. Ahimbisibwe, "Early risk prediction of gestational diabetes using routine antenatal care clinical data using machine learning," *Research Square Preprint*, 2026, doi: 10.21203/rs.3.rs-8721947/v1.
- [13] Y. Du et al., "An explainable machine learning-based clinical decision support system for prediction of gestational diabetes mellitus," *Scientific Reports*, vol. 12, no. 1170, 2022, doi: 10.1038/s41598-022-05112-2.
- [14] B. P. Pradeep Kumar, P. K. B. Rangaiah, and R. Augustine, "Comparative Analysis of Deep Learning Models for Diabetes Prediction," *SSRN Preprint*, 2025.

- [15] M. S. Al Reshan et al., “An Innovative Ensemble Deep Learning Clinical Decision Support System for Diabetes Prediction,” *IEEE Access*, vol. 12, pp. 106193-106210, 2024, doi: 10.1109/ACCESS.2024.3436641.
- [16] A. Y. Ahmed, “Personalized Glycemic Control Modeling Using Explainable Artificial Intelligence for Optimizing Diabetes Management,” *Scientia. Technology, Science and Society*, vol. 3, no. 1, pp. 85-96, 2026, doi: 10.59324/stss.2026.3(1).08.
- [17] Y.-N. Chan et al., “A machine learning approach for early prediction of gestational diabetes mellitus using elemental contents in fingernails,” *Scientific Reports*, vol. 13, no. 4184, 2023, doi: 10.1038/s41598-023-31270-y.
- [18] S. Sirmayanti, P. H. Prastyo, F. Rahman, and Mahyati, “Comparative Analysis of The Combination of Metaheuristic and Machine Learning Algorithms,” *IJID (International Journal on Informatics for Development)*, vol. 15, no. 1, pp. 672-682, 2026, doi: 10.14421/ijid.2026.4888.
- [19] A. El-Sayyid El-Bashbishy and H. M. El-Bakry, “Pediatric diabetes prediction using machine learning,” *Scientific Reports*, vol. 16, 2026, doi: 10.1038/s41598-025-24964-y.
- [20] R. Islam et al., “Clinical Decision Support System for Diabetic Patients by Predicting Type 2 Diabetes Using Machine Learning Algorithms,” *Journal of Healthcare Engineering*, vol. 2023, Art. ID 6992441, 2023, doi: 10.1155/2023/6992441.
- [21] P. Landa et al., *Operations Research and Artificial Intelligence in Healthcare Management*. Springer, 2026.
- [22] W. Peng, Z. Huang, and M. Zhang, “Genetic Evidence Links Gestational Diabetes Mellitus to Increased ER-Negative Breast Cancer Risk Through a Mendelian Randomization Analysis,” *International Journal of Endocrinology*, vol. 2026, Art. no. 7371037, 2026, doi: 10.1155/ije/7371037.