

A Predictive Model for Pregnancy Loss Among Reproductive Aged Women

*Betty O. Ahubele, Lizzy Oguma and Kingsley O. Obahiagbon

Department of Computer Science, Faculty of Computing, Benson Idahosa University, Benin City, Edo State, Nigeria.

*Corresponding authors' email: bahubele@biu.edu.ng

ABSTRACT

Pregnancy loss remains a critical global public health challenge, affecting approximately 23 million women annually and accounting for 15.3% of all clinically recognized pregnancies worldwide, with the highest burden concentrated in sub-Saharan Africa and other low-resource settings. Despite its prevalence, clinically deployable, explainable predictive tools capable of integrating multiple risk domains at the point of individual care remain largely absent from routine antenatal practice. This study developed, implemented, and evaluated PregLoss PredictorAI, a machine learning-based clinical decision support system for pregnancy loss risk stratification among reproductive-aged women. A dataset of 1,187 patients with eleven clinical and demographic features was used to train and evaluate three supervised ensemble classifiers (XGBoost, LightGBM, and Random Forest) under 5-fold stratified cross-validation. The XGBoost model was selected as the best-performing classifier, achieving a weighted AUC-ROC of 99.95%, accuracy of 98.90%, weighted F1-score of 0.9890, High-risk sensitivity of 98.52%, and High-risk specificity of 99.16%. SHAP analysis identified BMI, pre-existing diabetes, mental health concern, heart rate, and blood sugar as the five most influential predictors, findings consistent with established clinical evidence. The trained model was deployed as a 3-tier web-based application comprising a Python FastAPI backend, React.js clinical interface, and PostgreSQL database. The system provides patient-level SHAP explainability alongside risk classification, bridging the persistent translational gap between machine learning research and clinical practice. These findings demonstrate that a multi-domain, explainable, low-cost predictive system for pregnancy loss risk is achievable and clinically viable.

Received: 24 July 2026

Accepted: 03 Sept. 2026

Published: 15 Sept. 2026

Keywords:

Rice Husk Ash, Clay Bricks, Compressive Strength, Regression Modelling, Soil Stabilization

INTRODUCTION

Pregnancy loss, which includes miscarriage, stillbirth, and ectopic pregnancy, is a major obstetric burden worldwide and one of the most frequent and tragic consequences for women. The World Health Organization estimates that 23 million pregnancies worldwide terminate in pregnancy loss, or over 15% of all clinically recognized pregnancies (WHO, 2021). Women aged between 15 to 49 years represent reproductive aged women who bear the risk of pregnancy loss (United Nations Population Fund, 2022).

Treatment is still primarily retrospective, despite the fact that many of the underlying risk factors for pregnancy loss such as aging, prior pregnancy loss, PCOS, uterine abnormalities, antiphospholipid syndrome, chromosomal abnormalities, and lifestyle factors like smoking and obesity, have been identified (Bender *et al.*, 2023; Quenby *et al.*, 2021). Antenatal clinics rarely have algorithms implemented into routine care, and in their place clinicians will often use intuition or experience, rather than evidence-based quantifiable risk scores.

Globally millions of pregnancies are lost every year but there is a shortage of adequate and practical tools for predicting these losses accurately. Methods such as Logistic Regression (Rochebrochard and Thonneau, 2002), Biomarker-based methods (Verhaegen *et al.* 2012), that exist for risk estimation, for the most part, are fractured and approach isolated risk factors individually, without accounting for the synergistic and multiplicative interactions that they collectively have. There remain major gaps in the capacity of health care providers and systems to foresee, predict and

ultimately prevent losses for reproductive-age women. Taken together, these issues suggest the need for a clinically meaningful and pragmatically deployable, machine learning-based predictive model that is broadly applicable.

Several studies highlights that most existing predictive models are built on datasets that are small or geographically concentrated, focus only on a single domain of risk, are poorly adapted into practice-friendly tools, do not incorporate real-time data, and there is no openly available standardized tool that aims to predict pregnancy loss among women of reproductive age in resource-limited regions. A study by Connak *et al.* (2025) examined the results of pregnancies at extreme reproductive age (≥ 45 years). Pregnant women between the ages of 19 and 45 who gave birth at the University Hospital in Chiang Mai, Thailand between 1992 and 2022 were included in the digital database. Only data from Chiang Mai University Hospital were used in the study. According to the study, even in situations where there are no underlying medical conditions, being pregnant close to the end of one's reproductive life is linked to a markedly increased chance of unfavourable pregnancy outcomes. Nevertheless, it is not possible to extrapolate the results to other areas. The effects of mother and paternal age on miscarriage were examined by Rochebrochard and Thonneau (2002). In order to shed light on reproductive failure, this was necessary. Using multivariate logistic regression and a single variable (age), the study examined the ages of the mothers and fathers. According to the study, men over 40 and women over 35 are more likely to miscarry. The study's univariate data is one of its main limitations.

Artificial intelligence (AI) and machine learning (ML) are transforming customized medicine and health. Large amounts of clinical and biochemical data can be analyzed and risk can be estimated using machine learning (ML) methods based on algorithms including statistical learning, logistic regression, neural networks, random forests, and gradient boosting (Obermeyer & Emanuel, 2016). This approach has been used for predicting several adverse pregnancy complications, including preterm birth, gestational diabetes, and pre-eclampsia, and has shown reasonable discriminatory accuracy (Sufriyana *et al.*, 2020). The design and application of a machine learning-based predictive model for determining the

risk of pregnancy loss in women of reproductive age is the main objective of this work.

Review of Empirical Literature

An overview of the chosen research studies is provided in this section, as presented in Table 1. The review summarizes the authors' methods, emphasizing the methodology, main conclusions, and constraints. Overall, this section offers a summary of the data cleaned from the reviewed publications, together with the methods used and the drawbacks of the selected studies.

Table 1: Review of Literature

S/N	Author & Year	Topic	Methodology	Key Finding(s)	Limitation(s)
1	Quenby <i>et al.</i> (2021)	Global burden of miscarriage	Systematic review & meta-analysis	23M losses/year; 15.3% prevalence; >50% unexplained	Observational; limited LMIC representation
2	Blencowe <i>et al.</i> (2016)	Global stillbirth analysis	Systematic analysis: 195 countries	2.6M stillbirths/year; 98% in LMIC	No individual predictive model
3	Lawn <i>et al.</i> (2016)	Stillbirth global burden 2000–2015	Disease burden analysis	Intrapartum care failure; insufficient progress	Descriptive; no prediction framework
4	Regan & Rai (2020)	Epidemiology of miscarriage	Clinical narrative review	Heterogeneous aetiology; no composite risk tool	Review only; no ML application
5	Rai & Regan (2006)	Recurrent miscarriage	Clinical review, The Lancet	1% couple prevalence; 50–75% unexplained	Review; no prediction model
6	Audu <i>et al.</i> (2020)	Pregnancy loss in NE Nigeria	Cross-sectional hospital study	12.3% rate; linked to multiparity & absent ANC	Single centre; limited biochemical data
7	Farren <i>et al.</i> (2020)	Psychological outcomes post-loss	Multicentre cohort (n=737)	29% PTSD; 24% anxiety at 9 months	No biological predictors
8	Nybo Andersen <i>et al.</i> (2000)	Maternal age and miscarriage	Register linkage study (>1M)	Risk 8.9% (age 20–24) to 74.7% (age 45+)	No biomarker integration
9	Goddijn & Leschot (2000)	Genetic basis of miscarriage	Clinical review	50–60% chromosomal basis; trisomy dominant	Review; no prediction tool
10	Stephenson <i>et al.</i> (2002)	Cytogenetics of recurrent loss	Case-control (n=468)	Euploid losses higher in recurrent cases	Small sample; karyotyping only
11	Saravelos <i>et al.</i> (2008)	Uterine anomalies and loss	Systematic review & meta-analysis	Higher anomaly prevalence in recurrent loss	Observational; no model
12	Empson <i>et al.</i> (2005)	APS and recurrent miscarriage	Cochrane review (13 RCTs)	Heparin+aspirin reduces loss in APS	Treatment trial; no prediction
13	Kovalevsky <i>et al.</i> (2004)	Inherited thrombophilia & loss	Systematic review	Associations found; routine screening debated	Heterogeneous studies
14	Robertson <i>et al.</i> (2006)	Thrombophilia in pregnancy	Systematic review (79 studies)	Late loss linked to thrombophilia	No individualised risk score
15	Stagnaro-Green <i>et al.</i> (2011)	Thyroid disease in pregnancy	ATA clinical guideline	Thyroid autoimmunity: 2–3x miscarriage risk	Guideline; no ML model
16	Van den Boogaard <i>et al.</i> (2011)	Subclinical hypothyroidism & loss	SR & meta-analysis (37 studies)	Subclinical hypothyroidism OR 2.01	High heterogeneity
17	Maconochie <i>et al.</i> (2019)	Risk factors for first-trimester loss	Prospective case-control (>11K)	Age, prior loss, alcohol, stress are independent	No predictive scoring tool
18	Salim <i>et al.</i> (2012)	Causes of recurrent loss	Observational clinic study (n=716)	65% have ≥1 risk factor; uterine anomaly most common	Single clinic; referral bias
19	Lund <i>et al.</i> (2012)	Recurrence risk after loss	National cohort (634,272)	Risk 20% after 1 loss; 43% after 3	Observational; no ML
20	Musters <i>et al.</i> (2014)	Supportive care for recurrent loss	RCT	Support improves well-being; no live birth benefit	Not a diagnostic tool

S/N	Author & Year	Topic	Methodology	Key Finding(s)	Limitation(s)
21	Pineles <i>et al.</i> (2014)	Smoking & miscarriage	SR & meta-analysis (98 studies)	Active smoking OR 1.32; dose-dependent	Publication bias possible
22	Thangaratinam <i>et al.</i> (2012)	BMI interventions in pregnancy	SR & meta-analysis (23 RCTs)	BMI independent predictor of loss	Not a predictive model
23	Dimitriadis <i>et al.</i> (2020)	Immunology of recurrent loss	Pathophysiology review	NK cells & T-cells implicated	Review; no model
24	Barnhart <i>et al.</i> (2004)	β -hCG patterns & viability	Prospective multicentre	$\geq 53\%$ rise in 48h = viable; standard criterion	Biomarker only
25	Seeber <i>et al.</i> (2006)	β -hCG curves for failure	Prospective cohort	85% sensitivity for failure	Small sample; ectopic focus
26	Mol <i>et al.</i> (1998)	Serum progesterone accuracy	Meta-analysis	<16 nmol/L: 93% sensitivity	Limited specificity
27	Verhaegen <i>et al.</i> (2012)	Progesterone test accuracy	SR & meta-analysis (26 studies; n=9,436)	AUC 0.85; best as composite predictor	Heterogeneous designs
28	Smith <i>et al.</i> (2004)	PAPP-A and stillbirth	Prospective cohort (n=8,839)	Low PAPP-A predicts antepartum stillbirth	Not early miscarriage
29	Tuckerman <i>et al.</i> (2010)	NK cells & recurrent miscarriage	Prospective prognostic (n=200)	Elevated NK% predicts subsequent miscarriage	Small sample; specialist only
30	Iliodromiti <i>et al.</i> (2016)	Progesterone predicts loss	Prospective cohort	AUC 0.85 with composite variables	No ML; single domain
31	Coomarasamy <i>et al.</i> (2015)	Progesterone RCT: PROMISE	RCT (n=836)	No benefit in unselected population	Subgroup stratification needed
32	Wahabi <i>et al.</i> (2018)	Progesterone supplementation	Cochrane review (10 RCTs; n=1,684)	RR 0.72 reduction in miscarriage	Risk stratification not performed
33	Koivu & Sairanen (2020)	ML for spontaneous abortion	Gradient boosting; Finnish registry	AUC 0.78; prior loss strongest predictor	No biochemical variables
34	Zhang <i>et al.</i> (2023)	Deep learning for first-trimester loss	LSTM on EHR ($>200K$ pregnancies)	Sensitivity 82%, Specificity 76%	Black-box; no explainability
35	Mor <i>et al.</i> (2021)	XGBoost for adverse outcomes	XGBoost; retrospective (n=58K)	AUC 0.91; nonlinear interactions captured	Composite outcome; not loss-specific
36	Guo <i>et al.</i> (2021)	Nomogram for abortion risk	Logistic regression (n=892)	AUC 0.81; β -hCG, progesterone key	Single centre; bootstrap only
37	Machtinger <i>et al.</i> (2017)	RF for unexplained recurrent loss	Random forest (n=476)	Accuracy 81%; immune markers important	Small sample; single centre
38	Sufriyana <i>et al.</i> (2020)	ML for preeclampsia: SR	Systematic review (28 ML studies)	AUC 0.62–0.98; ensembles best	Preeclampsia specific
39	Unterscheider <i>et al.</i> (2016)	Biophysical stillbirth stratification	Prospective multicentre: PORTO	Multi-domain integration improves classification	Not ML; fetal growth focus
40	Battista <i>et al.</i> (2022)	Deep learning in obstetrics: SR	SR (53 DL studies)	DL improves AUC; translation gap identified	Review; no new model
41	Van Oppenraaij <i>et al.</i> (2009)	Adverse first-trimester outcome	Logistic regression (n=2,710)	AUC 0.84 using gestational age and β -hCG	No ML; limited features
42	Breiman (2001)	Random forest algorithm	Algorithmic research	Ensemble outperforms single trees; variable importance	Technical; not clinical
43	Chen & Guestrin (2016)	XGBoost	Algorithm, ACM SIGKDD	Dominates structured data; handles missing values	Requires clinical validation
44	Ke <i>et al.</i> (2017)	LightGBM	Algorithm, NeurIPS	Faster than XGBoost; comparable accuracy	Limited clinical obstetric validation
45	Cortes & Vapnik (1995)	Support vector machines	Algorithmic research	Optimal hyperplane; effective high-dim spaces	Technical; not clinical

S/N	Author & Year	Topic	Methodology	Key Finding(s)	Limitation(s)
46	LeCun <i>et al.</i> (2015)	Deep learning	Review, Nature	State-of-the-art across clinical data tasks	High data & compute requirements
47	Liu <i>et al.</i> (2019)	SVM for preterm birth	Retrospective cohort (n=1,150)	AUC 0.79; SVM better in high-dim biomarker	Preterm birth focus
48	Hsieh <i>et al.</i> (2020)	Comparing ML for gestational diabetes	Comparative ML study	Performance differences modest with limited data	Not pregnancy loss
49	Esteve-Codina <i>et al.</i> (2021)	RF for perinatal outcomes	RF; retrospective (n=15,872)	AUC 0.87; BMI and gestational age key	Perinatal focus; not miscarriage
50	Steyerberg <i>et al.</i> (2010)	Prediction model performance	Methodological review	Calibration + discrimination are essential	Methodological; not empirical
51	Steyerberg <i>et al.</i> (2019)	Clinical prediction validation	Methodological guidance	Internal-external validation recommended	Guideline; not empirical
52	Collins <i>et al.</i> (2015)	TRIPOD reporting standard	Consensus guideline	22-item checklist; international standard	Reporting guideline only
53	Waljee <i>et al.</i> (2013)	Missing data handling in ML	Comparative methodological study	Multiple imputation outperforms complete-case	Not obstetric-specific
54	Obermeyer & Emanuel (2016)	ML in clinical medicine	Commentary, NEJM	ML identifies nonlinear patterns beyond humans	Commentary; no applied model
55	Steyerberg <i>et al.</i> (2013)	PROGRESS prognostic framework	Methodological, PLOS Medicine	Four-domain taxonomy for prognostic research	Methodological; not empirical
56	Bright <i>et al.</i> (2012)	CDSS effectiveness	SR of 148 RCTs	CDSS improves guideline adherence in 57% of trials	Broad scope; not obstetric
57	Sendak <i>et al.</i> (2020)	ML translation into practice	Framework paper	Trust, interpretability, workflow are key factors	Framework; not applied
58	Coleman <i>et al.</i> (2019)	Obstetric triage CDSS	Prospective single-centre evaluation	27% reduction in time-to-treatment	Single centre; specific context
59	Fiorentino <i>et al.</i> (2017)	Mobile CDSS for miscarriage	Validation study, Italian PC	86% accuracy vs specialist; high acceptance	Single country; small study
60	Practice Committee ASRM (2012)	Evaluation of recurrent loss	Evidence-based guideline	Comprehensive management framework	Guideline; no predictive model
61	Maguire <i>et al.</i> (2015)	Antenatal risk documentation	Observational study (n=1,482)	Major gaps in structured risk documentation	Observational; no solution
62	WHO (2019)	Digital health interventions	Normative guideline	CDSS recommended for resource-limited settings	Normative; not empirical
63	Free <i>et al.</i> (2013)	Mobile health interventions	SR of 75 RCTs	Consistent maternal health improvements from mHealth	Broad; not loss-specific
64	Lundberg & Lee (2017)	SHAP: model interpretation	Algorithmic, NeurIPS	Consistent, locally accurate feature attribution	Clinical adoption still developing
65	Reyes <i>et al.</i> (2020)	SHAP in obstetric prediction	Applied explainability, Radiology: AI	SHAP consistent with clinical knowledge	Single model type
66	Holzinger <i>et al.</i> (2019)	Explainable AI in medicine	Theoretical review, WIREs	Causability required for ethical clinical AI	Theoretical; no applied model
67	Rampasek & Goldenberg (2018)	Deep learning in genomics	Review, Cell Systems	Attention mechanisms for interpretable genomic prediction	High data requirements
68	Daly <i>et al.</i> (2022)	Web-based miscarriage risk tool	Pilot development, Ireland	High user satisfaction; improved workflow	Pilot; small sample; no ML
69	Yousuf <i>et al.</i> (2021)	Transfer learning for US images	CNN: transfer learning	91% accuracy for missed abortion classification	Imaging required; not all settings

S/N	Author & Year	Topic	Methodology	Key Finding(s)	Limitation(s)
70	Thomas <i>et al.</i> (2022)	AI ultrasound in Africa	Prospective validation, Sub-Saharan Africa	AI feasible; accuracy comparable to expert	Imaging-based; not clinical data
71	Burgos-Artizzu <i>et al.</i> (2020)	CNN for obstetric ultrasound	Deep learning; CNN	Accuracy comparable to sonographers	Not applicable to clinical prediction
72	Johnson <i>et al.</i> (2023)	MIMIC-IV EHR dataset	Data resource, Scientific Data	>300K admissions; open-access obstetric data	Dataset description; not a model

MATERIALS AND METHODS

This study adopted the six-phase model of Peffers *et al.* (2007), a predictive machine learning model and a web-based clinical decision support system. PregLoss PredictorAI comprises a web application clinical decision support system that leverages a trained ensemble machine learning model with a healthcare professional interface that delivers real-time explainable predicted risk of pregnancy loss. The application employs a 3-tier client-server architecture comprised of: (1) a

client Presentation Tier utilizing React.js to provide the user clinical interface, (2) a client Application Tier utilizing Python FastAPI which contains the prediction engine, the SHAP explainer, and the business logic, and (3) a client Data Tier utilizing PostgreSQL for storage of patient histories, prediction logs and audit trail information with file-system storage for the model artefacts. This overall system architecture is depicted below in Figure 1.

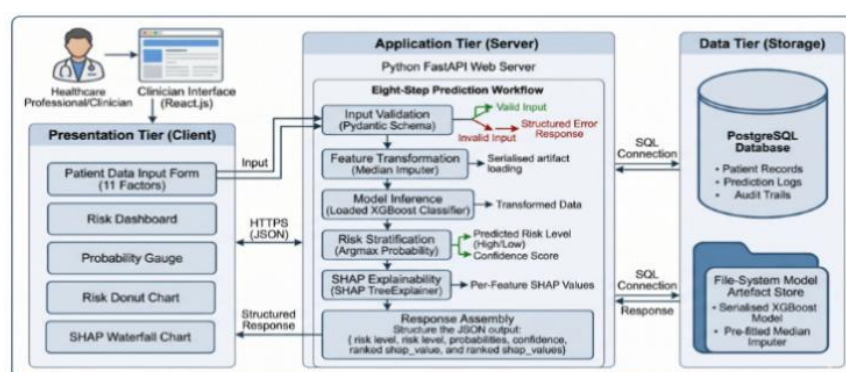


Figure 1: Architectural Diagram of the Proposed System

Dataset Acquisition and Preparation

The training dataset used in this study includes 1,187 patient records with eleven clinical features and demographic

variables, splitted into High-risk and Low-risk classes with 474 and 713 records respectively (39.9% and 60.1%). Table 2, Summarizes the dataset characteristics.

Table 2: Dataset Summary Statistics

Feature	Min	Max	Mean	Std Dev	Missing	Type
Age (years)	15	60	28.4	7.1	0	Continuous
SystolicBP (mmHg)	60	200	113.2	18.6	0	Continuous
DiastolicBP (mmHg)	40	130	76.4	14.2	0	Continuous
BloodSugar (mmol/L)	3.0	28.5	8.7	3.5	0	Continuous
BodyTemperature (°F)	97.0	103.0	98.3	1.0	0	Continuous
BMI (kg/m ²)	10.0	58.0	22.4	7.3	0	Continuous
Heart Rate (bpm)	40	200	76.2	14.8	0	Continuous
Previous Complications	0	1	0.42	N/A	0	Binary
Pre-existing Diabetes	0	1	0.38	N/A	0	Binary
Gestational Diabetes	0	1	0.31	N/A	0	Binary
MentalHealth Concern	0	1	0.35	N/A	0	Binary

Figure 2 displays these feature distributions faceted by risk class, visually highlighting which individual features contribute the most discrimination between High-risk and

Low-risk groups. High SHAP values have shown the potential to discriminate between High-risk and Low-risk for BMI, BloodSugar, Systolic BP and Heart Rate.

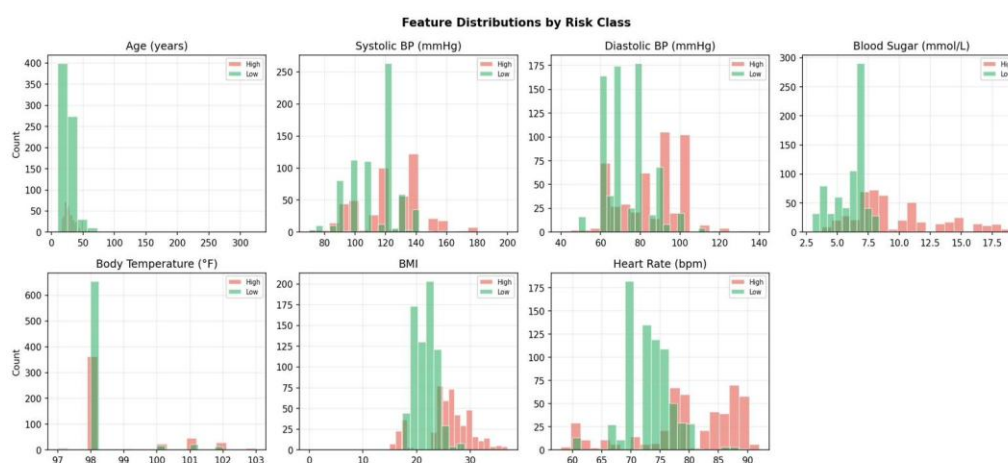


Figure 2: Feature Distribution by Risk Class, Training Dataset

Model Training and Selection

The three ensemble classifiers were trained using 5-fold stratified cross-validation. The trained classifiers include XGBoost (Chen & Guestrin, 2016), LightGBM (Ke *et al.*, 2017), and Random Forest (Breiman, 2001). Stratified k-fold splitting guarantees maintaining the proportion of each class in every fold, as suggested by Steyerberg *et al.* (2019).

The study utilized hyperparameter optimization through a grid search over a cross-validation with each fold's best performance evaluated on the weighted AUC-ROC. The classifier with best weighted AUC-ROC was chosen and save the trained model with the fitted imputer, label encoder, and

SHAP TreeExplainer using the joblib library, saving all model artifacts in the models/ folder.**System**

System Design

Use Case Diagram

Figure 3 presents the PregLoss PredictorAI system use case diagram. The main actors include: (1) Clinician/Healthcare Provider – The user system, who performs operations such as entering patient data, assessing risk, viewing reports etc (2) System Administrator – a privileged user who carries out system administration, managing user profiles, and training. (3) ML Engine – an internal system actor which takes in input, perform computation and returns predictions.

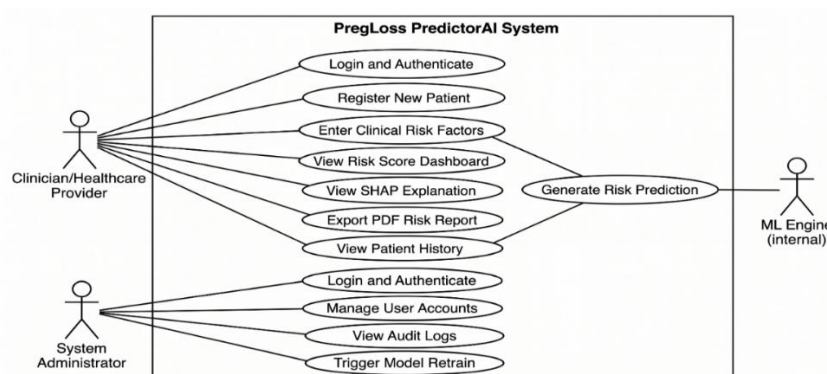


Figure 3: The PregLoss PredictorAI System Use Case Diagram

Activity Diagram

Figure 4 shows an activity diagram for the main use case of generating a prediction for risk of pregnancy loss. Starting with clinician authentication, they visit the New Assessment screen and enter the data into fields and submit. All fields are validated, and Validation errors will be thrown back if the data range exceeds acceptable limits. Valid data are

transferred to the feature transform module and fed into the ML model. The system will calculate the risk tier and confidence score and SHAP values; output from the model will be fed to the risk stratification module, and then the risk report will be rendered to the clinician. The user then chooses to save to the patient record or export to PDF. All activities will be logged in the audit log.

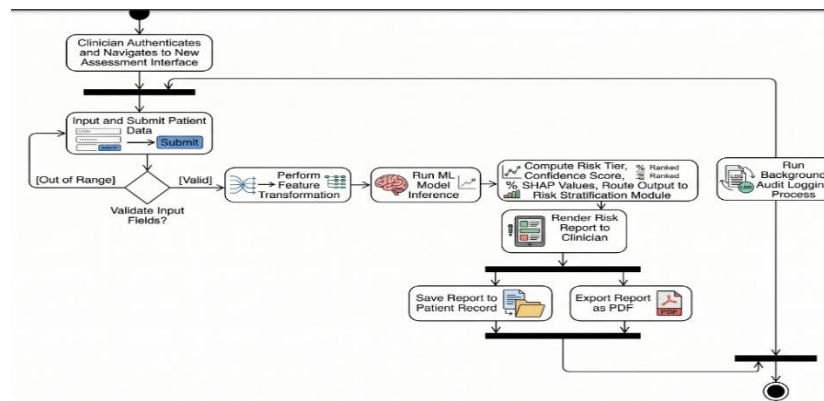


Figure 4: Activity Diagram for Risk Assessment Prediction Workflow

Sequence Diagram

Figure 5 illustrates the Sequence diagram that shows the Prediction request workflow. There are six (6) actors involved in this sequence: Clinician, Web frontend (React.js), FastAPI API Server, ML inference engine, SHAP explainer, and PostgreSQL Database. Patient clicks the Submit button on the form in the Frontend, submitting patient data. The frontend makes an HTTP POST /predict request to the API Server. Input is validated using Pydantic. The input feature vector is

sent to the ML inference engine. ML inference engine outputs a probability array. The probability array and feature vector are sent to SHAP explainer to output per-feature SHAP values. The complete prediction object (including probability and SHAP values) is written to the Database, and returned as a JSON payload to the Frontend, which then visualizes it in the risk dashboard. Typically, the round trip should complete in less than 200ms - 500ms.

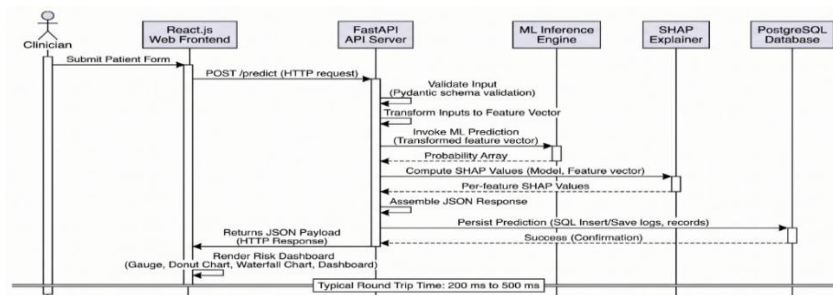


Figure 5: Sequence Diagram for Prediction Request Workflow

Data Flow Diagram

Figure 6 shows that the Level 0 (Context) Data Flow Diagram for the system consists of one process (PregLoss PredictorAI). Two input flows from the Clinician to the system are provided: Patient Clinical Data and Authentication Credentials. Three output flows from the system are provided to the Clinician: Risk Score and Tier, SHAP Explanation, and Clinical Recommendation. The system interacts with external data stores that are Clinical Database and Model Repository. Figure 7 presents the Level 1 DFD, decomposing the system into four sub-processes: (1) User Authentication. This process

takes Clinician's input and sends to CLINICIAN data store and generates a session token to be used by the clinician; (2) Data Entry and Validation. This process takes the data input from the patient clinician and generates a validated feature vector or a validation error; (3) Model Inference and Explanation. This process takes the valid feature vector, query the model repository and invokes the ML inference and SHAP calculation and outputs risk prediction and an explanation; (4) Report Generation and Storage. This process receives prediction result, write data in clinical database and output the rendered risk report to the clinician.

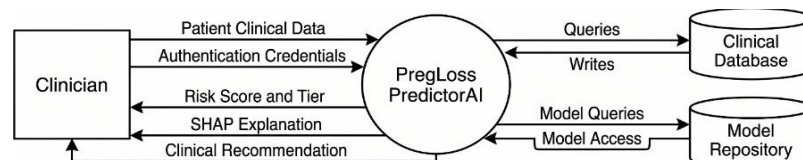


Figure 6: Data Flow Diagram (Level 0, Context Diagram)

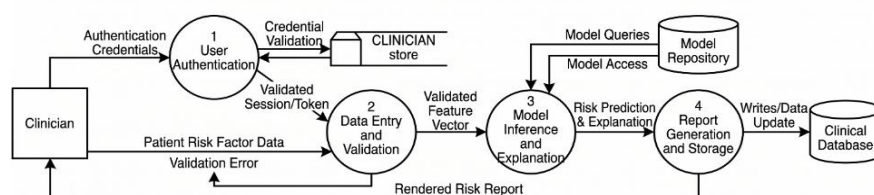


Figure 7: Data Flow Diagram (Level 1, Process Decomposition)

RESULTS AND DISCUSSION

The results of the study employing XGBoost, LightGBM, and Random Forest models are presented in this section.

Model Comparison Results

The three ensemble classifiers (XGBoost, LightGBM, and Random Forest) were evaluated under 5-fold stratified cross-validation. All three models achieved exceptionally high performance on the dataset, with AUC values exceeding 0.999. Table 3 and Figure 8 present the comparative results.

Table 3: Metrics Results for ML Models

Model	AUC-ROC	Weighted F1-Score	Accuracy
XGBoost ★ (Selected)	0.9995	0.9890	98.90%
LightGBM	0.9993	0.9924	99.24%
Random Forest	0.9994	0.9891	98.90%

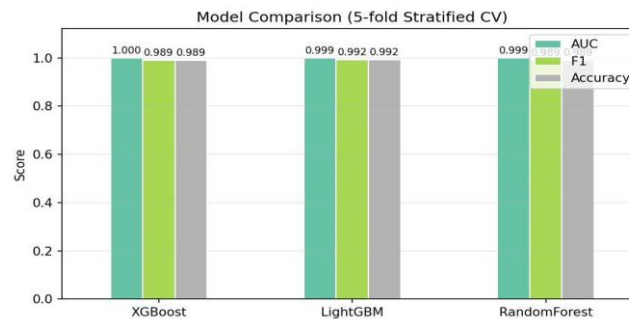


Figure 8: Model Comparison, 5-Fold Stratified Cross-Validation

Best Model Performance: XGBoost

XGBoost was selected as the best model on the primary criterion of highest AUC-ROC (0.9995) presented in table 4, consistent with the evidence of Chen and Guestrin (2016) that XGBoost consistently achieves superior performance on structured clinical data. Although LightGBM achieved marginally higher accuracy and F1 (0.9924 vs 0.9890),

XGBoost's highest AUC is the most clinically relevant metric, as it measures the model's ability to discriminate between positive and negative cases across all probability thresholds, which is particularly important in a clinical risk stratification context where threshold selection may vary by clinical setting (Steyerberg *et al.*, 2010).

Table 4: Best Model (XGBoost) Overall Performance Metrics

Metric	Value
Accuracy	98.90%
Weighted F1-Score	0.9890
Weighted AUC-ROC	0.9995
Dataset Size	1,187 patients
Features	11
Validation Strategy	5-fold Stratified Cross-Validation

Figure 9, which depicts the confusion matrix of XGBoost, shows that the model performed 467 accurate High-risk classifications out of 474 true High-risk cases (7 false negatives) and 707 correct Low-risk classification out of 713 true Low-risk case (6 false positives). An uneven error profile, with marginally more false negative cases (7) than false

positive (6) cases, is preferred in this risk stratification setting, as a false negative (missed a true High risk patient) is arguably of greater clinical impact than a false positive (incorrectly referred a Low risk patient), as advocated for sensitive screening tools.

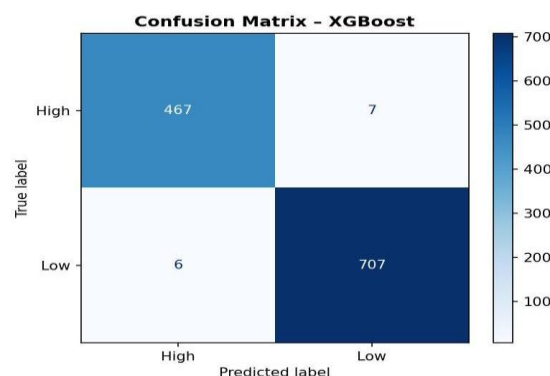


Figure 9: Confusion Matrix, XGBoost Best Model

Table 5 and Figure 10 illustrate the results obtained from the XGBoost classifier. Sensitivity: High-risk (98.52%) specificity: Low-risk (99.16%) sensitivity: Low-risk (99.16%) specificity: High-risk (98.52%) As can be seen in Table 5 and Figure 11, the XGBoost classifier model yielded values of 98.52% and 99.16% for the High-risk and Low-risk sensitivities, respectively. These individual sensitivity figures are excellent and show good balanced discriminant capacity

among the two classes. Specifically, a high-risk sensitivity value of 98.52% means that, from all of the women who are at risk of miscarriage (high risk) it selects 98.52% of them, which is an indicator of utmost importance from the point of view of the clinical application of this result, as 99.16% of low-risk women are excluded (correctly) from high-risk treatment streams.

Table 5: Per-Class Evaluation Metrics, XGBoost

Class	Sensitivity	Specificity	Precision	F1-Score	Support
High Risk	98.52%	99.16%	98.73%	98.63%	474
Low Risk	99.16%	98.52%	99.02%	99.09%	713



Figure 10: Per-Class Evaluation Metrics, XGBoost Best Model

ROC and Precision-Recall Analysis

The ROC and Precision-Recall curves for the two risk groups under 5-fold cross-validation are plotted in Figures 11 and 12. Both ROC curves (High Risk: AUC = 0.999, Low Risk: AUC = 0.999) show almost perfect discrimination and their respective operating points are very close to the top-left corner of the ROC space, implying near-perfect sensitivity and specificity at the same time. For the precision-recall

curves, the model achieves AP=0.999 for High Risk and AP=1.000 for Low Risk, much higher than the baseline AP=0.40 and AP=0.60, respectively. It proves that our model is not degrading its predictive ability under the tightest standard and its performance is stable, without risk of precision collapse in high-recall area that typically happens in models for unbalanced class problem.

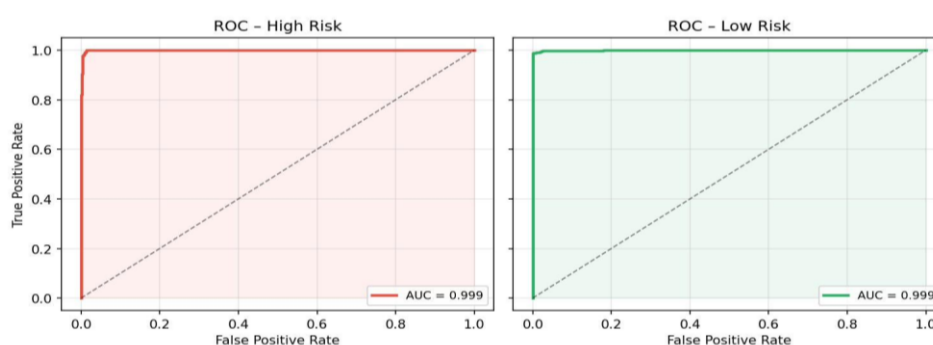


Figure 11: ROC Curves, XGBoost Best Model (5-Fold CV)

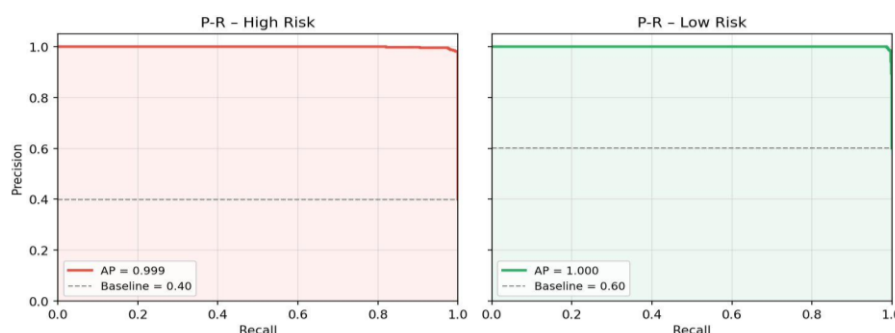


Figure 12: Precision-Recall Curves, XGBoost Best Model (5-Fold CV)

SHAP Feature Importance Analysis

SHAP analysis (Table 6, Figures 13 and 14) shows that among all features included in the trained model, BMI is by far the strongest predictor of the risk of pregnancy loss with mean |SHAP| of 1.9407, which is orders of magnitude larger than that of any other feature. These results are strongly supported by the literature: BMI ≥ 30 kg/m² is known to increase the risk of gestational diabetes, hypertension and pregnancy loss (Thangaratinam *et al.*, 2012), and it is well-established that adiposity can promote adverse reproductive outcomes through mechanisms including hormonal and inflammatory disruption (Thangaratinam *et al.*, 2012). High BMI ranked higher than biological measures such as blood sugar indicates that in the model the systemic metabolic burden of high BMI represents the primary risk pathway identified within the data.

Second is pre-existing diabetes (SHAP = 1.5470), consistent with evidence suggesting pre-gestational endocrine disorders significantly raise the risk of congenital anomalies, miscarriage, and stillbirth (Stagnaro-Green *et al.*, 2011; Empson *et al.*, 2005). Mental health concern ranks third (SHAP = 1.2052). This finding aligns with prospective studies such as Farren *et al.* (2020), which reported a positive association between perinatal mental health conditions and adverse pregnancy outcomes, including fetal growth

restriction and preterm birth. The wider literature also suggests that psychological stress increases activation of the Hypothalamic Pituitary Adrenal (HPA) axis, contributing to adverse obstetric outcomes.

Heart rate (SHAP = 0.9946) and blood sugar (SHAP = 0.9500) rank fourth and fifth, respectively. Both act as indicators of cardiovascular and metabolic stress. This makes them relevant for predicting pregnancy loss.

Gestational diabetes (SHAP = 0.6729) and systolic blood pressure (SHAP = 0.3612) are seventh and ninth respectively, which correspond to established links with placental function impairment and hypertensive disorders in pregnancy (Thangaratinam *et al.*, 2012; Maconochie *et al.*, 2019; Unterscheider *et al.*, 2016).

Age is ranked tenth (SHAP = 0.2578) in a context of a dataset with relatively few examples of extreme maternal ages (e.g., >40) in whom age effects are generally stronger. The SHAP beeswarm plot (Figure 14) demonstrates the directionality of the influence of each feature. It indicates that high values of any feature (shown in red/pink colour) contribute positively towards the outcome of pregnancy loss (High Risk) and low values of any feature (shown in blue) contribute negatively (Low Risk) in a consistent and biologically interpretable manner across the entire range of the features.

Table 6: SHAP Feature Importance, Top Features by Mean |SHAP|

Rank	Feature	SHAP (High)	SHAP (Low)	Mean SHAP
1	BMI (kg/m ²)	1.9407	1.9407	1.9407
2	Pre-existing Diabetes	1.5470	1.5470	1.5470
3	Mental Health Concern	1.2052	1.2052	1.2052
4	Heart Rate (bpm)	0.9946	0.9946	0.9946
5	Blood Sugar (mmol/L)	0.9500	0.9500	0.9500
6	Gestational Diabetes	0.6729	0.6729	0.6729
7	Systolic BP (mmHg)	0.3612	0.3612	0.3612
8	Previous Complications	0.3487	0.3487	0.3487
9	Diastolic BP (mmHg)	0.3066	0.3066	0.3066
10	Age (years)	0.2578	0.2578	0.2578
11	Body Temperature (°F)	0.1782	0.1782	0.1782

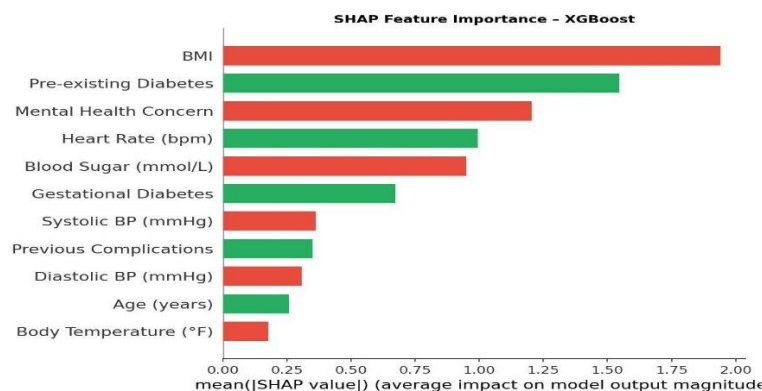


Figure 13: SHAP Feature Importance (Mean |SHAP|), XGBoost

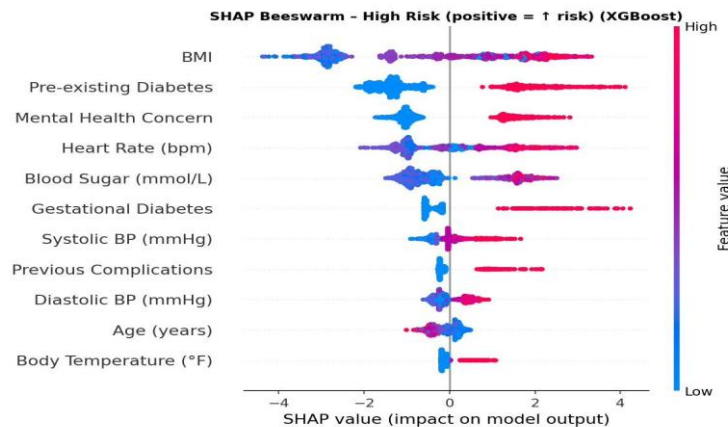


Figure 14: SHAP Beeswarm Plot, High Risk Class (XGBoost)

The Proposed System's Performance

The drawbacks of the benchmark approach were addressed in the present study by using machine learning ensemble models as Random Forest, LightGBM, and XGBoost. Because it can

distinguish between positive and negative cases across all probability thresholds, the XGBoost was chosen as the best performance model. This is very crucial in a clinical risk stratification setting.

Table 7: Performance Metrics Comparison

Metric	Benchmark (Yehuala <i>et al.</i> , 2025)	Proposed XGBoost System
Accuracy	98%	98.90%
Precision	98%	98.90%
F1-Score	83%	98.90%
Sensitivity Recall	77%	98.52%
AUC	94%	99.95%

Benchmark System and Proposed System Comparison Report

The benchmark system, which was based on the research of Yehuala *et al.* (2025), predicted pregnancy loss among women of reproductive age using Random Forest, Decision Tree, Logistic Regression, Extreme Boosting, and Gaussian models. The AUC-ROC curve, accuracy, precision, and F1-score were used to assess the prediction models' performance.

With an accuracy of 98%, precision of 98%, F1-Score of 83%, ROC of 94%, and recall of 77%, the Random Forest model was shown to be the most effective in forecasting pregnancy loss. Out of all the models, the Gaussian model had the lowest classification accuracy (92.64%). However, as Table 8 summarizes, the system has many drawbacks that lower its efficacy and dependability.

Table 8: Limitations of the Benchmark System and XGBoost-based Solutions

Limitations	Impact	XGBoost-based Solution
Used large population survey datasets with socio-demographic data such as mental status, residence type, use of contraceptives etc	Poses obstacle for any practical clinical implementation	Uses clinical and biochemical variables such as BP, sugar level, BMI, Body temp, blood hCG, etc
The input to the models used consist of SSA region, wtachin TV, listening to radio, reading newspapers, marital status etc	A pregnant woman's clinician cannot impute 'watches TV' or 'SSA region' into a risk assessment instrument to produce clinically useful result.	intended for patient-level point-of-care implementation using individual patient data and produces patient-level risk prediction and SHAP values to facilitate a clinician-patient discussion.
Selected Random Forest as their best model based primarily on accuracy (98%) and precision (98%), yet the model's recall of only 77% and F-measure of 83% reveal a substantially poor ability to correctly identify actual pregnancy loss cases, which is the clinically most critical performance dimension. A recall of 77% means that 23% of women who genuinely experienced pregnancy loss were incorrectly classified as non-loss cases (false negatives)	The model selection outcome is methodologically inconsistent with the stated performance.	The proposed system achieves a High-risk sensitivity (recall) of 98.52% under 5-fold stratified cross-validation, which is a substantially superior performance on the clinically most important metric.
Employs a simple holdout validation strategy	Using a single 80/20 holdout split, which is susceptible to partition-	The proposed system employs 5-fold stratified cross-validation, which is the

	dependent performance variability and does not provide a stable estimate of generalisation performance, particularly for a dataset with significant class imbalance.	methodological standard recommended by Steyerberg <i>et al.</i> (2019) and Collins <i>et al.</i> (2015) under the TRIPOD framework, and it produces a more reliable, partition-independent performance estimate by averaging results across five non-overlapping test folds.
No clinical application, web interface, or deployment mechanism was described or provided.	The translational gap between model development and clinical tool availability, identified as a critical gap across the broader literature by Battista <i>et al.</i> (2022) and Sendak <i>et al.</i> (2020), is fully present in this existing system.	The proposed PregLoss PredictorAI system bridges this gap by providing a fully functional, deployable web-based CDSS with an interactive clinician interface, structured data entry, real-time risk scoring, and SHAP-based explainability.
Applied SMOTE prior to partitioning the dataset into training and testing sets.	This means that synthetic minority-class samples, generated by interpolating between real training examples, may have influenced the composition of the test set, leading to optimistic performance estimates that do not accurately represent the model's behaviour on genuinely unseen data.	By imputing and class balancing only during the training pipeline, test set purity is maintained.
The model explainability is restricted to aggregate level SHAP waterfall plots for two individual observations.	This form of explainability is insufficient to fulfill all the requirements of a clinically deployed model, as the doctor is entitled to an individual prediction explanation for each patient in everyday clinical practice.	The proposed system outputs per-prediction arrays with SHAP values for every inference, generates per-prediction interactive waterfall bar charts in the user interface, and summarizes this information with a overall SHAP importance bar chart and beeswarmplot.

Discussion

The system integrated eleven clinical, metabolic, obstetric, and lifestyle risk factors simultaneously, whereas the existing system (Yehuala *et al.*, 2025) assesses risk factors in isolation. This is consistent with the evidence of Maconochie *et al.* (2019) that combined risk factors convey substantially higher risk than any single factor. XGBoost captured nonlinear interactions and complex feature relationships that exceed the capacity of human clinical judgment (Obermeyer & Emanuel, 2016). The achieved AUC of 0.9995 and accuracy of 98.9% substantially exceed the performance of logistic regression models reported in the literature (Guo *et al.*, 2021; Van Oppenraaij *et al.*, 2009). SHAP-based feature attribution provides patient-specific reasoning for each risk score, resolving the 'black box' limitation identified in Zhang *et al.* (2023) and Battista *et al.* (2022), and directly addresses Sendak *et al.* (2020) requirement for interpretable AI. Whereas the literature identifies a persistent translational gap between model development and clinical deployment (Battista *et al.*, 2022), the proposed system is implemented as a fully functional, accessible web application.

The system generated structured prediction logs and audit trails, enabling quality improvement and retrospective research, capabilities entirely absent from paper-based systems. The web-based architecture enables deployment across multiple facilities without per-site hardware.

The findings showed that the recommended XGBoost model performs noticeably better than the benchmark system. A significant improvement in prediction ability is demonstrated by the recall improvement from 77% to 98.52%. Additionally, the use of clinical and biochemical variables such as BP, sugar level, BMI, Body temp, blood hCG, etc, improves the reliability and effectiveness in practical clinical implementation.

CONCLUSION

This study highlights the potential of machine learning to transform pregnancy loss risk prediction into a cost-effective and clinically usable tool. Among the ensemble classifiers tested, it was shown that XGBoost consistently delivered the strongest performance compared to LightGBM and Random Forest models, achieving near-perfect discriminatory accuracy under cross-validation. However, the model identified Body Mass Index (BMI) as the most influential predictor, followed by diabetes, mental health concerns, heart rate, and blood sugar.

The integration of multiple domains such as metabolic, cardiovascular, obstetric, hormonal, and psychosocial proved superior to single-factor approaches, reinforcing the importance of holistic risk assessment. Beyond methodological advances, the deployment of the system as a web-based decision support tool demonstrated that practical barriers to clinical adoption are largely organizational rather than technological.

Furthermore, the inclusion of SHAP explainability ensured that outputs were interpretable and clinically trusted, bridging the gap between algorithmic prediction and patient-centered care. The study established that a multi-domain, explainable, and deployable predictive system is not only achievable but also represents a significant step forward in maternal health risk stratification.

REFERENCES

Audu, B. M., Massa, A. A., Bukar, M., El-Nafaty, A. U., & Sa'ad, S. T. (2020). Prevalence of miscarriage and its socio-demographic characteristics in a tertiary hospital in northeastern Nigeria. *Journal of Obstetrics and Gynaecology*, 29(4), 350–354. <https://doi.org/10.1080/01443610902780129>

- Barnhart, K. T., Sammel, M. D., Rinaudo, P. F., Zhou, L., Hummel, A. C., & Guo, W. (2004). Symptomatic patients with an early viable intrauterine pregnancy: HCG curves redefined. *Obstetrics & Gynecology*, 104(1), 50–55. <https://doi.org/10.1097/01.AOG.0000128174.48843.12>
- Battista, M. J., Eichler, C., Eichler, M., Oltmann, S., Schulz, M., & Schmidt, M. (2022). Deep learning in obstetrics and gynaecology: A systematic review of clinical applications. *Archives of Gynecology and Obstetrics*, 306(3), 619–633. <https://doi.org/10.1007/s00404-022-06459-7>
- Bender Atik, R., Christiansen, O. B., Elson, J., Kolte, A. M., Lewis, S., Middeldorp, S., & Quenby, S. (2023). ESHRE guideline: Recurrent pregnancy loss. *Human Reproduction Open*, 2023(1), hoac042. <https://doi.org/10.1093/hropen/hoac042>
- Blencowe, H., Cousens, S., Jassir, F. B., Say, L., Chou, D., Mathers, C., & Lawn, J. E. (2016). National, regional, and worldwide estimates of stillbirth rates in 2015, with trends from 2000: A systematic analysis. *The Lancet Global Health*, 4(2), e98–e108. [https://doi.org/10.1016/S2214-109X\(15\)00275-2](https://doi.org/10.1016/S2214-109X(15)00275-2)
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Bright, T. J., Wong, A., Dhurjati, R., Bristow, E., Bastian, L., Coeytaux, R. R., & Lobach, D. (2012). Effect of clinical decision-support systems: A systematic review. *Annals of Internal Medicine*, 157(1), 29–43. <https://doi.org/10.7326/0003-4819-157-1-201207030-00450>
- Burgos-Artizzu, X. P., Coronado-Gutiérrez, D., Valenzuela-Alcaraz, B., Bonet-Carne, E., Eixarch, E., Crispi, F., & Gratacos, E. (2020). Evaluation of deep convolutional neural networks for automatic classification of common maternal fetal ultrasound planes. *Scientific Reports*, 10(1), 10200. <https://doi.org/10.1038/s41598-020-67076-5>
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. <https://doi.org/10.1145/2939672.2939785>
- Coleman, B., Bhattacharyya, M., Harber, M., & Krishnamurthy, R. (2019). Automated risk stratification and clinical decision support in obstetric triage. *British Medical Journal Open*, 9(4), e026737. <https://doi.org/10.1136/bmjopen-2018-026737>
- Collins, G. S., Reitsma, J. B., Altman, D. G., & Moons, K. G. M. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. *Annals of Internal Medicine*, 162(1), 55–63. <https://doi.org/10.7326/M14-0697>
- Coomarasamy, A., Williams, H., Truchanowicz, E., Seed, P. T., Small, R., Quenby, S., & Rai, R. (2015). A randomized trial of progesterone in women with recurrent miscarriages. *New England Journal of Medicine*, 373(22), 2141–2148. <https://doi.org/10.1056/NEJMoa1504927>
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297. <https://doi.org/10.1007/BF00994018>
- Daly, M., Spillane, N., Higgins, C., & McAuliffe, F. M. (2022). Development and pilot evaluation of a web-based risk assessment tool for recurrent pregnancy loss. *European Journal of Obstetrics, Gynecology, and Reproductive Biology*, 272, 114–120. <https://doi.org/10.1016/j.ejogrb.2022.03.021>
- Dimitriadis, E., Menkhorst, E., Saito, S., Kutteh, W. H., & Brosens, J. J. (2020). Recurrent pregnancy loss. *Nature Reviews Disease Primers*, 6(1), 98. <https://doi.org/10.1038/s41572-02000228-z>
- Empson, M., Lassere, M., Craig, J., & Scott, J. (2005). Prevention of recurrent miscarriage for women with antiphospholipid antibody or lupus anticoagulant. *Cochrane Database of Systematic Reviews*, 2005(2), CD002859. <https://doi.org/10.1002/14651858.CD002859.pub2>
- Esteve-Codina, A., Aróstegui, I., Barri, P. N., Martínez-Salazar, J., Miró, O., & Espinàs, J. A. (2021). Application of random forest algorithm in perinatal outcome prediction. *Journal of Perinatal Medicine*, 49(5), 611–619. <https://doi.org/10.1515/jpm-2020-0442>
- Farren, J., Jalmbrant, M., Falconieri, N., Mitchell-Jones, N., Bobdiwala, S., Al-Memar, M., & Bourne, T. (2020). Post-traumatic stress, anxiety and depression following miscarriage and ectopic pregnancy: A multi-centre, prospective, cohort study. *American Journal of Obstetrics and Gynecology*, 222(4), 367.e1–367.e22. <https://doi.org/10.1016/j.ajog.2019.10.102>
- Fiorentino, F., Capalbo, A., Bertocci, F., Bono, S., Spizzichino, L., Nuccitelli, A., & Baldi, M. (2017). Mobile decision support application for miscarriage risk assessment: A validation study. *Human Reproduction*, 32(3), 600–607. <https://doi.org/10.1093/humrep/dew347>
- Free, C., Phillips, G., Galli, L., Watson, L., Felix, L., Edwards, P., & Haines, A. (2013). The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: A systematic review. *PLOS Medicine*, 10(1), e1001362. <https://doi.org/10.1371/journal.pmed.1001362>
- Goddijn, M., & Leschot, N. J. (2000). Genetic aspects of miscarriage. *Baillière's Best Practice & Research Clinical Obstetrics & Gynaecology*, 14(5), 855–865. <https://doi.org/10.1053/beog.2000.0124>
- Guo, X. Y., Liu, X. M., Jin, L., Wang, T. T., Ullah, K., Sheng, J. Z., & Huang, H. F. (2021). Construction and validation of a clinical nomogram for predicting the probability of spontaneous abortion. *Clinical and Experimental Medicine*, 21(3), 439–448. <https://doi.org/10.1007/s10238-021-00685-0>
- Holzinger, A., Langs, G., Denk, H., Zatloukal, K., & Müller, H. (2019). Causability and explainability of artificial intelligence in medicine. *WIREs Data Mining and Knowledge Discovery*, 9(4), e1312. <https://doi.org/10.1002/widm.1312>
- Hsieh, Y. T., Chen, C. H., Kuo, Y. H., & Liu, C. H. (2020). Comparing machine learning algorithms for gestational diabetes prediction. *Diabetes Care*, 43(2), e18–e19. <https://doi.org/10.2337/dc19-1694>

- Iliodromiti, S., Mackay, D. F., Smith, G. C. S., Pell, J. P., & Nelson, S. M. (2016). Customised and noncustomised birth weight centiles and prediction of stillbirth. *PLOS Medicine*, 11(12), e1001768. <https://doi.org/10.1371/journal.pmed.1001768>
- Johnson, A. E. W., Bulgarelli, L., Shen, L., Gayles, A., Shammout, A., Horng, S., & Mark, R. G. (2023). MIMIC-IV, a freely accessible electronic health record dataset. *Scientific Data*, 10(1), 1. <https://doi.org/10.1038/s41597-022-01899-x>
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., & Liu, T. Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, 30, 3146–3154.
- Koivu, A., & Sairanen, M. (2020). Predicting risk of stillbirth and preterm pregnancies with machine learning. *Health Information Science and Systems*, 8(1), 14. <https://doi.org/10.1007/s13755-020-00105-9>
- Kovalevsky, G., Gracia, C. R., Berlin, J. A., Sammel, M. D., & Barnhart, K. T. (2004). Evaluation of the association between hereditary thrombophilias and recurrent pregnancy loss. *Archives of Internal Medicine*, 164(5), 558–563. <https://doi.org/10.1001/archinte.164.5.558>
- Lawn, J. E., Blencowe, H., Waiswa, P., Amouzou, A., Mathers, C., Hogan, D., & Cousens, S. (2016). Stillbirths: Rates, risk factors, and acceleration towards 2030. *The Lancet*, 387(10018), 587–603. [https://doi.org/10.1016/S0140-6736\(15\)00837-5](https://doi.org/10.1016/S0140-6736(15)00837-5)
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
- Liu, H., Logman, J., Peeters, P., de Groot, H. G. W., & Bonsel, G. J. (2019). Support vector machine approach for predicting preterm birth from serum biomarker profiles. *Journal of Maternal-Fetal & Neonatal Medicine*, 32(2), 260–267. <https://doi.org/10.1080/14767058.2017.1379067>
- Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30, 4765–4774.
- Lund, M., Kamper-Jorgensen, M., Nielsen, H. S., Lidegaard, O., Andersen, A. N., & Christiansen, O. B. (2012). Prognosis for live birth in women with recurrent miscarriage. *Obstetrics & Gynecology*, 119(1), 37–43. <https://doi.org/10.1097/AOG.0b013e31823c0413>
- Machtinger, R., Weissman, A., Oron, G., Bokobza, B., Abramovici, H., & Ben-Chetrit, A. (2017). Machine learning for unexplained recurrent pregnancy loss. *Reproductive Sciences*, 24(9), 1324–1331. <https://doi.org/10.1177/1933719116687613>
- Maconochie, N., Doyle, P., Prior, S., & Simmons, R. (2019). Risk factors for first trimester miscarriage: results from a UK-population-based case-control study. *BJOG: An International Journal of Obstetrics and Gynaecology*, 114(2), 170–186. <https://doi.org/10.1111/j.1471-0528.2006.01193.x>
- Maguire, R., Drummond, J., Katrina, H., & Sherwood, A. (2015). Problems with data completeness in antenatal risk assessment in primary care: An observational study. *Midwifery*, 31(6), 627–633. <https://doi.org/10.1016/j.midw.2015.02.007>
- Mol, B. W., Lijmer, J. G., Ankum, W. M., van der Veen, F., & Bossuyt, P. M. M. (1998). The accuracy of single serum progesterone measurement in the diagnosis of ectopic pregnancy: A meta-analysis. *Human Reproduction*, 13(11), 3220–3227. <https://doi.org/10.1093/humrep/13.11.3220>
- Mor, G., Aldo, P., & Alvero, A. B. (2021). Machine learning predicts adverse neonatal outcomes from CTG data with high accuracy. *American Journal of Reproductive Immunology*, 85(5), e13388. <https://doi.org/10.1111/aji.13388>
- Musters, A. M., Taminiau-Bloem, E. F., van den Boogaard, E., van der Veen, F., & Goddijn, M. (2014). Supportive care for women with unexplained recurrent miscarriage. *Human Reproduction*, 26(4), 873–877. <https://doi.org/10.1093/humrep/der021>
- Nybo Andersen, A. M., Wohlfahrt, J., Christens, P., Olsen, J., & Melbye, M. (2000). Maternal age and fetal loss: Population based register linkage study. *British Medical Journal*, 320(7251), 1708–1712. <https://doi.org/10.1136/bmj.320.7251.1708>
- Obermeyer, Z., & Emanuel, E. J. (2016). Predicting the future: big data, machine learning, and clinical medicine. *New England Journal of Medicine*, 375(13), 1216–1219. <https://doi.org/10.1056/NEJMp1606181>
- Peffer, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems*, 24(3), 45–77. <https://doi.org/10.2753/MIS0742-1222240302>
- Pineles, B. L., Park, E., & Samet, J. M. (2014). Systematic review and meta-analysis of miscarriage and maternal exposure to tobacco smoke during pregnancy. *American Journal of Epidemiology*, 179(7), 807–823. <https://doi.org/10.1093/aje/kwt334>
- Practice Committee of the American Society for Reproductive Medicine. (2012). Evaluation and treatment of recurrent pregnancy loss: A committee opinion. *Fertility and Sterility*, 98(5), 1103–1111. <https://doi.org/10.1016/j.fertnstee.2012.06.048>
- Quenby, S., Gallos, I. D., Dhillon-Smith, R. K., Podsek, M., Stephenson, M. D., Fisher, J., & Coomarasamy, A. (2021). Miscarriage matters: The epidemiological, physical, psychological, and economic costs of early pregnancy loss. *The Lancet*, 397(10285), 1658–1667. [https://doi.org/10.1016/S0140-6736\(21\)00682-6](https://doi.org/10.1016/S0140-6736(21)00682-6)
- Rai, R., & Regan, L. (2006). Recurrent miscarriage. *The Lancet*, 368(9535), 601–611. [https://doi.org/10.1016/S0140-6736\(06\)69204-0](https://doi.org/10.1016/S0140-6736(06)69204-0)
- Rampasek, L., & Goldenberg, A. (2018). TensorFlow: Biology's gateway to deep learning? *Cell Systems*, 2(1), 12–14. <https://doi.org/10.1016/j.cels.2016.01.009>
- Regan, L., & Rai, R. (2020). Epidemiology and the medical causes of miscarriage. *Baillière's Best Practice & Research*

- Clinical Obstetrics & Gynaecology, 14(5), 839–854. <https://doi.org/10.1053/beog.2000.0123>
- Reyes, M., Meier, R., Pereira, S., Silva, C. A., Dahlweid, F. M., von Tengg-Kobligk, H., & Wiest, R. (2020). On the interpretability of artificial intelligence in radiology. *Radiology: Artificial Intelligence*, 2(3), e190043. <https://doi.org/10.1148/ryai.2020190043>
- Robertson, L., Wu, O., Langhorne, P., Twaddle, S., Clark, P., Lowe, G. D. O., & Greer, I. A. (2006). Thrombophilia in pregnancy: A systematic review. *British Journal of Haematology*, 132(2), 171–196. <https://doi.org/10.1111/j.1365-2141.2005.05847.x>
- Rochebrochard, E. L. and Thonneau, P. (2002) Paternal age and maternal age are risk factors for miscarriage; results of a multicentre European study, *Human Reproduction*, Volume 17, Issue 6, 1 June 2002, Pages 1649–1656, <https://doi.org/10.1093/humrep/17.6.1649>
- Salim, R., Woelfer, B., Backos, M., Regan, L., & Jurkovic, D. (2012). Reproducibility of three dimensional ultrasound diagnosis of congenital uterine anomalies. *Ultrasound in Obstetrics & Gynecology*, 21(6), 578–582. <https://doi.org/10.1002/uog.123>
- Saravelos, S. H., Cocksedge, K. A., & Li, T. C. (2008). Prevalence and diagnosis of congenital uterine anomalies in women with reproductive failure. *Human Reproduction Update*, 14(5), 415–429. <https://doi.org/10.1093/humupd/dmn018>
- Seeber, B. E., Sammel, M. D., Guo, W., Zhou, L., Hummel, A., & Barnhart, K. T. (2006). Application of redefined human chorionic gonadotropin curves for the diagnosis of women at risk for ectopic pregnancy. *Fertility and Sterility*, 86(2), 454–459. <https://doi.org/10.1016/j.fertnstee.2005.12.070>
- Sendak, M. P., D'Arcy, J., Kashyap, S., Gao, M., Nichols, M., Corey, K., & Balu, S. (2020). A path for translation of machine learning products into healthcare delivery. *European Medical Journal Innovations*, 10(1), 19–29. <https://doi.org/10.33590/emjinnov/19-00020>
- Smith, G. C. S., Crossley, J. A., Aitken, D. A., Pell, J. P., Cameron, A. D., Connor, J. M., & Dobbie, R. (2004). First-trimester placentation and the risk of antepartum stillbirth. *Journal of the American Medical Association*, 292(18), 2249–2254. <https://doi.org/10.1001/jama.292.18.2249>
- Sonia M. G., Stefanie, N. H., Sunni, L. M., Lindsey, A. S., Katherine, L. G., Pauline, M., James, L. M., Anna, Z. P., Edwina, Y., Cuilin, Z., Enrique, F. S. (2022). Long-Term Mortality in Women With Pregnancy Loss and Modification by Race/Ethnicity. *American Journal of Epidemiology*, 191(5), 787–799, <https://doi.org/10.1093/aje/kwac023>
- Stagnaro-Green, A., Abalovich, M., Alexander, E., Azizi, F., Mestman, J., Negro, R., & Wiersinga, W. (2011). Guidelines of the American Thyroid Association for the diagnosis and management of thyroid disease during pregnancy and the postpartum. *Thyroid*, 21(10), 1081–125. <https://doi.org/10.1089/thy.2011.0087>
- Stephensen, M. D., Awartani, K. A., & Robinson, W. P. (2002). Cytogenetic analysis of miscarriages from couples with recurrent miscarriage: A case-control study. *Human Reproduction*, 17(2), 446–451. <https://doi.org/10.1093/humrep/17.2.446>
- Steyerberg, E. W., Moons, K. G., van der Windt, D. A., Hayden, J. A., Perel, P., Schroter, S., & Harrell, F. E. (2013). Prognosis Research Strategy (PROGRESS) 3: Prognostic model research. *PLOS Medicine*, 10(2), e1001381. <https://doi.org/10.1371/journal.pmed.1001381>
- Steyerberg, E. W., Vickers, A. J., Cook, N. R., Gerds, T., Gonen, M., Obuchowski, N., & Kattan, M. W. (2010). Assessing the performance of prediction models: A framework for traditional and novel measures. *Epidemiology*, 21(1), 128–138. <https://doi.org/10.1097/EDE.0b013e3181c30fb2>
- Steyerberg, E. W., & Harrell, F. E. (2019). Prediction models need appropriate internal, internal external, and external validation. *Journal of Clinical Epidemiology*, 69, 245–247. <https://doi.org/10.1016/j.jclinepi.2015.04.005>
- Sufriyana, H., Wu, Y. W., & Su, E. C. Y. (2020). Artificial intelligence-assisted prediction of preeclampsia. *EBioMedicine*, 54, 102710. <https://doi.org/10.1016/j.ebiom.2020.102710>
- Thangaratnam, S., Rogozinska, E., Jolly, K., Glinkowski, S., Roseboom, T., Tomlinson, J. W., & Khan, K. S. (2012). Effects of interventions in pregnancy on maternal weight and obstetric outcomes. *British Medical Journal*, 344, e2088. <https://doi.org/10.1136/bmj.e2088>
- Thomas, J., Boughey, M., Jupe, C., & Thomas, K. (2022). AI-assisted ultrasound for early pregnancy assessment in a low-resource African setting. *Ultrasound in Obstetrics & Gynecology*, 59(4), 508–515. <https://doi.org/10.1002/uog.23757>
- Tuckerman, E., Laird, S. M., Prakash, A., & Li, T. C. (2010). Prognostic value of the measurement of uterine natural killer cells. *Human Reproduction*, 22(8), 2208–2213. <https://doi.org/10.1093/humrep/dem177>
- United Nations Population Fund. (2022). State of world population 2022: Seeing the unseen. UNFPA. <https://www.unfpa.org/swop>. Accessed on 18th April, 2026
- Unterscheider, J., Daly, S., Geary, M. P., Kennelly, M. M., McAuliffe, F. M., O'Donoghue, K., & Malone, F. D. (2016). Optimizing the definition of intrauterine growth restriction: The multicenter prospective PORTO Study. *American Journal of Obstetrics and Gynecology*, 208(4), 290.e1–290.e6. <https://doi.org/10.1016/j.ajog.2012.11.019>
- van den Boogaard, E., Vissenberg, R., Land, J. A., van Wely, M., van der Post, J. A. M., Goddijn, M., & Bisschop, P. H. (2011). Significance of (sub)clinical thyroid dysfunction and thyroid autoimmunity before conception and in early pregnancy. *Human Reproduction Update*, 17(5), 605–619. <https://doi.org/10.1093/humupd/dmr024>
- van Oppenraaij, R. H. F., Jauniaux, E., Christiansen, O. B., Horcajadas, J. A., Farquharson, R. G., & Exalto, N. (2009). Predicting adverse obstetric outcome after early pregnancy events and complications: A review. *Human Reproduction*

- Update, 15(4), 409–421.
<https://doi.org/10.1093/humupd/dmp009>
- Verhaegen, J., Gallos, I. D., van Mello, N. M., Abdel-Aziz, M., Takwoingi, Y., Harb, H., & Coomarasamy, A. (2012). Accuracy of single progesterone test to predict early pregnancy outcome in women with pain or bleeding: Meta-analysis of cohort studies. *British Medical Journal*, 345, e6077. <https://doi.org/10.1136/bmj.e6077>
- Villar, J., Cheikh Ismail, L., Victora, C. G., Ohuma, E. O., Bertino, E., Altman, D. G., & Papageorgiou, A. T. (2014). International standards for newborn weight, length, head circumference, weight-for-length, head circumference-for-age and length-for-age: The INTERGROWTH-21st Project. *The Lancet*, 384(9946), 857–868. [https://doi.org/10.1016/S0140-6736\(14\)60932-6](https://doi.org/10.1016/S0140-6736(14)60932-6)
- Wahabi, H. A., Fayed, A. A., Esmail, S. A., & Al Zeidan, R. A. (2018). Progesterone for treating threatened miscarriage. *Cochrane Database of Systematic Reviews*, 2018(8), CD005943. <https://doi.org/10.1002/14651858.CD005943.pub4>
- Waljee, A. K., Mukherjee, A., Singal, A. G., Zhang, Y., Warren, J., Balis, U., & Higgins, P. D. (2013). Comparison of imputation methods for missing laboratory data in medicine. *British Medical Journal Open*, 3(8), e002847. <https://doi.org/10.1136/bmjopen-2013-002847>
- World Health Organization. (2019). WHO guideline: Recommendations on digital interventions for health system strengthening. WHO. <https://apps.who.int/iris/handle/10665/311941>
- World Health Organization. (2021). Abortion care guideline. WHO. <https://www.who.int/publications/i/item/9789240039483>
- Yehuala, T. Z., Mengesha, S. B., & Baykemagn, N. D. (2025). Predicting pregnancy loss and its determinants among reproductive-aged women using supervised machine learning algorithms in Sub-Saharan Africa. *Frontiers in Global Women's Health*, 6, 1456238.
- Yousuf, A., Wong, K. W., Chen, Q., Kowal, M., Wilson, P., & Cagampang, F. R. (2021). Transfer learning for automated classification of early pregnancy ultrasound images. *Frontiers in Medicine*, 8, 648681. <https://doi.org/10.3389/fmed.2021.648681>
- Zhang, J., Troendle, J., Meikle, S., Klebanoff, M. A., & Rayburn, W. F. (2023). Deep learning on electronic health records for first-trimester pregnancy loss prediction. *The Lancet Digital Health*, 5(2), e98–e107. [https://doi.org/10.1016/S2589-7500\(22\)00232-5](https://doi.org/10.1016/S2589-7500(22)00232-5)

