1.0 Introduction

Maternal health remains one of the major public health problems for low- and middle-income countries (LMICs)[1]. In these regions, the impact of preeclampsia and/or eclampsia is felt even more severely, contributing significantly to maternal and neonatal morbidity and mortality [2]. Preeclampsia, typically marked by new-onset hypertension and proteinuria after 20 weeks of gestation[3], often manifests in severe, unpredictable forms requiring early clinical detection. Globally, it is directly responsible for 70,000 maternal fatalities annually, making it the second most common cause of maternal death [4]. In Sub-Saharan Africa, the overall pooled incidence of preeclampsia is 13% [5], exceeding the global average of 5–8% [6]. Despite this burden, antenatal care utilisation remains low, partly due to cultural beliefs that prior safe deliveries ensure future ones, leading to late detection of complications like preeclampsia [7]. According to [5], the severity of preeclampsia-related mortality can be decreased by prompt escalation to hospital-based treatment and early detection during prenatal care. To address these gaps, we developed **MamaPEARL (Preeclampsia Early Alert and Response Lab)**, an AI-powered web and mobile system designed for low-resource African settings. MamaPEARL integrates continuous home-based monitoring via a wearable strap for vital signs and lab test results as key inputs. It parses this data through a trained machine learning model to predict risk levels, issue timely alerts to users, and connect high-risk women to the nearest doctor subscribed to the application, sharing a structured summary of their clinical interactions and health data for coordinated care.

2.0 Methodology

2.1 Dataset Description

The dataset, sourced from a study by [3], includes 11,006 expectant mothers who received prenatal care between 2005 and 2017 at Yonsei University Healthcare Centre in Seoul, Korea. It excludes cases of termination before 24 weeks due to miscarriage, fetal death, or early-onset preeclampsia, and deliveries outside the centre. Collected features span demographics (age, height, weight, BP, gestational age), medical and obstetric history (e.g., hypertension, diabetes), medications, and biochemical lab values, including BUN, creatinine, UPCR, uric acid, liver enzymes, lipid profile, haemoglobin, and glucose.

2.2 AI Approach

2.2.1 Exploratory Data Analysis

We conducted exploratory data analysis on our data to understand the underlying patterns, assess data quality, and guide preprocessing and modelling decisions.

Class Distribution

Compared to the control group, there were comparatively fewer instances of preeclampsia cases, indicating a significant imbalance [Figure 1] in the class groups. However, this imbalance necessitated the use of balanced algorithms and evaluation metrics sensitive to minority class performance.

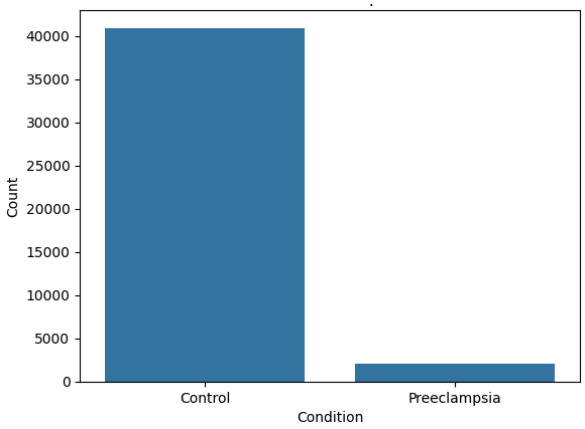

Figure 1: A count plot showing the class distribution. The control group (TGT = 0) significantly outnumbers the preeclampsia group (TGT = 1), indicating a class imbalance in the dataset.

Distributional Analysis of Key Clinical Features

Based on feature importance scores and supporting literature [3], we plotted histograms of selected clinical variables to examine their distribution by preeclampsia outcome. Representative examples are shown in Figure 2.

2.2.2. Addressing Class Imbalance

To avoid biased learning, we used a hybrid resampling technique because the original dataset showed a notable class imbalance (majority: around 13,000; minority: 300). In particular, we used random undersampling to downsample the majority class to 800 instances and SMOTE-based oversampling to upsample the minority class to 600 instances.

2.2.3. Model Training and Evaluation

We evaluated seven supervised learning algorithms: Random Forest, XGBoost, Gradient Boosting, Logistic Regression, Support Vector Machine, Naive Bayes, and Decision Tree. Each model was trained on the resampled dataset. Feature scaling was applied using standardisation, with .fit_transform() on training data and .transform() on test data. Hyperparameter tuning was conducted using GridSearchCV with 5-fold stratified cross-validation, optimising for average

precision score to reflect the importance of correctly identifying positive (preeclampsia) cases in an imbalanced setting.

2.2.4 Performance Metrics

Models were evaluated on the test set using multiple metrics:

- Balanced Accuracy: to ensure equal weight to both classes
- PR-AUC (Precision-Recall AUC): to quantify performance in skewed data
- Confusion Matrix: to understand class-wise prediction breakdown

2.3 Results and Findings

Among the models evaluated, Random Forest outperformed all others with a balanced accuracy of 81% and a PR-AUC of 89%, followed by Xtreme Boosting (78%, 83%) and Logistic Regression (71%, 80%). K-nearest Neighbours also demonstrated competitive performance. Decision Tree and Naive Bayes showed the weakest predictive power, particularly in terms of PR-AUC. These results highlight the advantage of ensemble methods in handling imbalanced clinical datasets. Full performance metrics are summarised in Table 1.

Model	**Balanced Accuracy**	**PR-AUC**
Decision Tree	66%	0.50
Naive Bayes	62%	0.60
Support Vector Classifier	69%	0.68
K-nearest Neighbours	72%	0.77
Logistic Regression	71%	0.80
Stochastic Gradient Boosting	71%	0.80
Xtreme Boosting	78%	0.83
Random Forest	**81%**	**0.89**

Table 1: Performance comparison of machine learning models for preeclampsia prediction on the resampled dataset.

2.4 System Architecture and Deployment

The MamaPEARL system was deployed as a mobile-responsive web application with a React frontend hosted on Vercel, a FastAPI backend on Render, and a Supabase database. The AI agent, built with Langchain and powered by GPT-4, enables multimodal interaction; users can either upload lab results (image input) or chat directly with the system (text input) to receive tailored maternal health support. Upon data submission, the backend parses clinical parameters via an NLP module and routes them to the supervised model trained. The system then returns real-time risk assessments, delivers preventive guidance, and escalates care to an in-app physician if necessary. This architecture, illustrated in Figure 3, enables continuous surveillance and timely intervention, especially in low-resource settings.

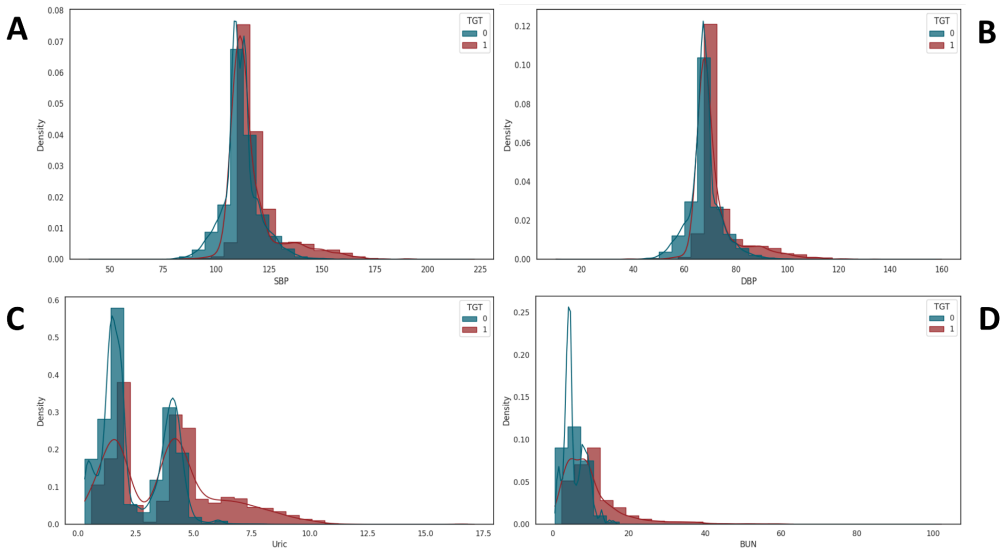

Figure 2. Distributions of selected clinical markers categorised by preeclampsia status. (a) Systolic blood pressure (SBP) values were more concentrated around 115 mmHg in the preeclampsia group, while the control group showed a broader spread. (b) Diastolic blood pressure (DBP) also demonstrated a rightward shift among preeclampsia cases. (c) Uric acid levels were elevated in the preeclampsia group, with a distribution extending into higher ranges. (d) Blood urea nitrogen (BUN) similarly showed higher values among women who developed preeclampsia, reflecting early renal involvement.

3.0 Discussion, Challenges, and Limitations

In sub-Saharan Africa, many pregnant women are discouraged from attending antenatal care due to long wait times, lack of subsidised tests and medications, or the belief that it's unnecessary after prior uncomplicated births. MamaPEARL was designed to ease this burden by integrating into

women's routines and enabling early detection of maternal risks like preeclampsia, which often goes undiagnosed due to delayed care and limited data.

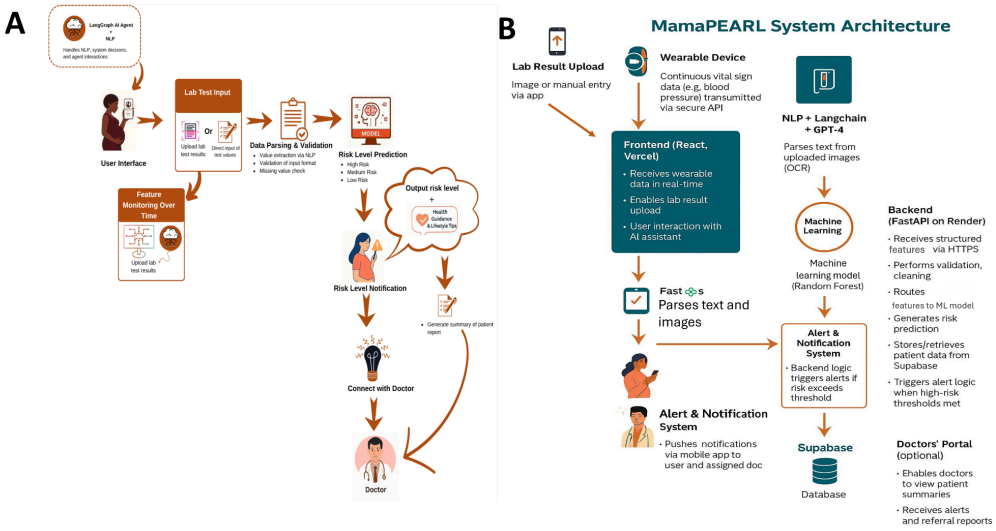

Figure 3: A: MamaPEARL System Architecture. B) MamaPEARL Process Flow. Application link.

Although our dataset included only participants enrolled after the first trimester, the Random Forest model showed strong predictive performance when combining maternal characteristics with second and third-trimester lab data. While class imbalance was initially a challenge, we addressed it using a hybrid sampling strategy. On the technical side, we encountered early deployment challenges related to digitising lab results using OCR. This issue was resolved by utilising the GPT-4 API for text recognition from scanned reports. Despite these constraints, our results support the integration of AI-driven tools into antenatal workflows to improve risk screening and maternal outcomes in low-resource settings.

4.0 Recommendation and Next Steps

We plan to transition from a web application to a fully native mobile application tailored for expectant mothers. To support continuous monitoring, we will integrate the MamaPEARL app with wearable straps that automatically transmit vital signs throughout pregnancy. A reminder feature will prompt scheduled lab tests, ensuring timely updates of clinical data from early pregnancy through delivery. Future development will focus on adding multilingual support for local languages and dialects, building a dedicated portal for healthcare providers to track patient alerts, and securing partnerships with hospitals and health agencies to drive adoption.

References

[1] K. R. Fox, J. C. Franklin, J. D. Ribeiro, E. M. Kleiman, K. H. Bentley, and M. K. Nock, "Meta-analysis of risk factors for nonsuicidal self-injury," Clin. Psychol. Rev., vol. 42, pp. 156–167, Dec. 2015, doi: 10.1016/j.cpr.2015.09.002.1.0 Introduction

[2] A. Shah et al., "Cesarean delivery outcomes from the WHO global survey on maternal and perinatal health in Africa," Int. J. Gynecol. Obstet., vol. 107, no. 3, pp. 191–197, Dec. 2009, doi: 10.1016/j.ijgo.2009.08.013.

[3] J. H. Jhee et al., "Prediction model development of late-onset preeclampsia using machine learning-based methods," PLOS ONE, vol. 14, no. 8, p. e0221202, Aug. 2019, doi: 10.1371/journal.pone.0221202.

[4] R. Townsend, O'Brien ,Patrick, and A. and Khalil, "Current best practice in the management of hypertensive disorders in pregnancy," Integr. Blood Press. Control, vol. 9, pp. 79–94, Jul. 2016, doi: 10.2147/IBPC.S77344.

[5] B. L. Manandhar, V. Chongstuvivatwong, and A. Geater, "Antenatal Care And Severe Pre-Eclampsia In Kathmandu Valley," J. Chitwan Med. Coll., vol. 3, no. 4, Art. no. 4, 2013.

[6] B. Jikamo, M. Adefris, T. Azale, and K. Alemu, "Incidence, trends and risk factors of preeclampsia in sub-Saharan Africa: a systematic review and meta-analysis," PAMJ-One Health, vol. 11, no. 1, Art. no. 1, May 2023, doi: 10.11604/pamj-oh.2023.11.1.39297.

[7] A. Nakimuli et al., "Pregnancy, parturition and preeclampsia in women of African ancestry," Am. J. Obstet. Gynecol., vol. 210, no. 6, pp. 510-520.e1, Jun. 2014, doi: 10.1016/j.ajog.2013.10.879.