IJARST

International Journal of Advanced Research in Science and Technology

https://doi.org/10.62226/ijarst20262788

ISSN 2319 – 1783 (Print)

ISSN 2320 – 1126 (Online)

ADVANCING AI-DRIVEN EARLY WARNING SYSTEMS FOR PREDICTION AND PREVENTION OF POSTPARTUM HAEMORRHAGE

Ohia Paul Ndudiri and Olojido Joseph Bamikole

Department of information Technology, University of Port Harcourt, Rivers state Nigeria and Department of computer science Rufus Giwa polytechnic Owo ondo state Nigeria

Email: *Paul.ohia@uniport.edu.ng*

ARTICLEINFO

Article history:

Received	22 Aug 2026
Accepted	27 Aug 2026
Available online	31 Aug 2026

Keywords:

Postpartum haemorrhage; machine learning; early warning system; Random Forest; SHAP; maternal health..

Indexed in:

and in major libraries

ABSTRACT

Postpartum haemorrhage (PPH) remains one of the leading causes of maternal morbidity and mortality worldwide, particularly in low-resource healthcare settings where delayed diagnosis and limited access to timely intervention significantly increase adverse maternal outcomes. This study developed and evaluated an artificial intelligence (AI)-based early warning model for predicting PPH risk using machine learning techniques to support timely clinical decision-making. A retrospective dataset comprising 223 anonymised maternal records obtained from Mpilo Central Hospital, Zimbabwe, was analysed. Data preprocessing involved handling missing values, capping outliers, and engineering clinically relevant features including labour duration, binary risk indicators, and a composite risk score followed by Min-Max normalization. Two supervised machine learning algorithms, Random Forest (RF) and Multilayer Perceptron (MLP), were trained and evaluated using an 80:20 train-test split. The RF model demonstrated superior predictive performance, achieving an accuracy of 86.67%, precision of 89.47%, recall of 80.95%, F1-score of 85.00%, and an area under the receiver operating characteristic curve (ROC-AUC) of 0.8730. Model interpretability was enhanced using SHapley Additive exPlanations (SHAP), which identified labour duration and mode of delivery as the most influential predictors of PPH. Compared with previously reported machine learning approaches, the proposed framework achieves competitive predictive performance while maintaining transparency on a relatively small, real-world dataset from a resource-constrained setting. The model offers a practical decision support tool for the early identification of women at high risk of PPH, enabling prompt clinical intervention and potentially reducing PPH related maternal morbidity and mortality.

1. Introduction

Postpartum haemorrhage (PPH) is a major obstetric emergency and one of the leading causes of maternal morbidity and mortality worldwide. It is characterized by excessive bleeding following delivery that is typically abrupt and rapid in onset. Despite advances in obstetric care, PPH remains a major cause of preventable maternal mortality, particularly in low and middle-income countries. Its unpredictability and rapid progression underscore the need for early identification and timely treatment to improve survival and prevent serious complications.

Conventional risk assessment typically relies on clinical judgement and established risk factors. While practical, these approaches often fail to capture complex interactions between demographic, obstetric, and intrapartum variables, and scoring systems may not keep pace with real-time changes during labour. Artificial intelligence (AI) and machine learning (ML), which can operate on multidimensional data and detect patterns that conventional statistics may miss, offer a more responsive alternative. AI-based early warning systems can support data-driven, proactive decision-making in obstetric care.

Recent studies highlight the growing role of ML in PPH prediction. Raman et al. (2025) evaluated logistic regression, random forest, gradient boosting, and neural network models using demographic, historical, and dynamic labour data; gradient boosting and neural networks outperformed traditional approaches, with the neural network achieving 93% accuracy and an AUC-ROC of 0.96, aided by real-time labour information. Yao et al. (2025) developed five ML models on 1,225 vaginal births in Shanghai, with XGBoost performing best after variable selection; SHAP analysis identified midwife experience, fear of childbirth, parity, second-stage labour duration, episiotomy, and companionship as the most influential predictors. Xiao et al. (2025) applied convolutional neural networks to T2 weighted MRI scans of 280 pregnant women, where a combined deep learning and clinical model achieved an AUC of 0.868, compared with 0.834 for imaging alone and 0.784 for clinical variables alone, illustrating the benefit of multimodal data fusion. Lérias-Cambeiro et al. (2025) applied K-means clustering for feature selection followed by logistic regression, ridge regression, and random forest, with ridge logistic regression performing best (sensitivity 0.857, specificity 0.875, accuracy 0.871, AUC 0.907), indicating that hybrid statistical ML approaches can be highly robust.

This study aims to (i) design an AI-based early warning system for postpartum haemorrhage, (ii) implement and optimise machine learning models within the proposed framework, and (iii) assess the clinical relevance and predictive performance of the developed system.

Although the studies above demonstrate strong predictive performance, most rely on large, high-income country datasets, imaging infrastructure, or dynamic intrapartum monitoring systems that are rarely available in African obstetric units, and few report model interpretability alongside predictive accuracy. There remains a shortage of integrated, interpretable, and clinically deployable early-warning models that have been validated on real-world obstetric data from resource-constrained African settings. This study addresses this gap by developing and validating an interpretable machine learning framework combining Random Forest and MLP models with SHAP based explanations using routinely collected obstetric records from a Zimbabwean hospital, with an explicit focus on transparency and deployment readiness rather than predictive accuracy alone.

2. Materials and Methods

2.1 Materials

This research used both software and hardware resources to support data preprocessing, model development, training, evaluation, and interpretation. All computations were performed on an MSI laptop with an Intel Core i7-8750H processor, 16 GB RAM, and a 4 GB NVIDIA GeForce GTX 1060 GPU, which provided sufficient processing power for neural network training with optional GPU acceleration.

The software environment was managed with Anaconda to maintain a reproducible Python 3.11 environment, with all development and experimentation conducted in Jupyter Notebook. The implementation relied on the following open-source libraries: Pandas and NumPy for data loading, manipulation, and numerical computation; Scikit learn for preprocessing (imputation, encoding, scaling, train test splitting), the Random Forest classifier, and performance evaluation; PyTorch for building, training, and evaluating the Multilayer Perceptron Matplotlib and Seaborn for visualisation; SHAP for model interpretability; and Joblib for serialising the trained model and preprocessing components.

The dataset used was a publicly available, retrospective set of 223 anonymised patient records from Mpilo Central Hospital, Zimbabwe, obtained from Mendeley Data (https://data.mendeley.com/datasets/k7z2yywdn5/1) and provided in Microsoft Excel format. No additional physical materials or clinical equipment were required.

2.2 Methods

2.2.1 Data Description and Preprocessing

The dataset comprised 223 patient records and 16 variables covering demographic, obstetric, and neonatal information: maternal age, gestational age, gravida, parity, number of children alive, previous obstetric complications, obstetric abnormalities, labour duration, delivery method, estimated blood loss, perineal outcome, birth weight, sex of the newborn, HIV status, booking status, and patient identifier. These are standard obstetric records collected during routine clinical care and anonymised prior to analysis. PPH was defined, following standard clinical criteria, as blood loss of 1,000 mL or more within 24 hours of delivery.

Initial inspection showed the dataset to be largely complete, with only 12 missing values across four variables (labour duration: 3; perineal outcome: 2 birth weight: 2; sex of baby: 5). Descriptive statistics showed a mean maternal age of 28.14 years (SD = 6.98, range 15-44), mean gravida of 2.74, mean parity of 1.63, and mean estimated blood loss of 971.62 mL (range 500-3,000 mL), indicating the presence of some extreme values. Categorical fields such as previous obstetric complications also showed inconsistent text formatting (e.g., mixed case and free-text entries) that required standardisation.

Missing values were addressed in two stages: forward-fill was first applied to resolve short consecutive gaps, after which any remaining missing values in numerical columns were imputed using K-Nearest Neighbours imputation (k = 5). Outliers in numerical variables were treated by capping values at the 1st and 99th percentiles, which limited the influence of extreme observations such as the

maximum recorded blood loss of 3,000 mL without materially distorting the overall data distribution. The patient identifier and HIV status were excluded from outlier treatment as they are not continuous clinical measures.

2.2.2 Feature Engineering and Selection

Labour duration, originally recorded as free text (e.g., "8h 40min"), was converted to a continuous numeric variable in minutes using a custom parsing function (e.g., "8h 40min" to 520 minutes). Categorical obstetric history fields were converted into binary indicators, including a Has_Previous_C_S flag and additional binary variables for obstetric complications and abnormalities. A composite Risk_Score variable was engineered by combining clinically relevant binary indicators previous caesarean section, obstetric complications, abnormalities, gravida greater than one, maternal age below 18 years, and maternal age above 35 years - to capture cumulative obstetric risk in a single feature. Birth weight values were cleaned by removing textual units (e.g., "2800g" to 2800.0) and converting to a numeric scale. Categorical variables including delivery method (Abortion, Forceps, LSCS, NVD), perineal outcome, and sex of the newborn were label encoded.

Of the 16 original variables, three were excluded prior to modelling. The patient identifier was removed as a non-informative label carrying no predictive value. Estimated blood loss was excluded from the feature set because it was used to derive the binary PPH outcome itself; retaining it as a predictor would constitute data leakage and artificially inflate performance. HIV status, although clinically relevant to coagulopathy and PPH risk in African obstetric populations, was withheld from the current model because the field contained a high proportion of inconsistently documented and undisclosed entries, a common limitation of routinely collected records for a sensitive health status. Imputing or encoding this variable under such conditions - on a dataset of only 223 records - risked introducing bias rather than genuine predictive signal. HIV status is therefore flagged as a priority variable for inclusion in future work using larger, more completely documented datasets rather than omitted on grounds of irrelevance.

This left 13 predictor variables for modelling: maternal age, gestational age, gravida, parity, number of children alive, labour duration (minutes), birth weight, the engineered binary complication/abnormality indicators, the composite risk score, and the encoded delivery method, perineal outcome, and sex-of-baby variables. Features were retained on the basis of three criteria: (i) established clinical relevance to PPH risk in the obstetric literature, (ii) availability and completeness in routinely collected intrapartum records typical of resource-constrained settings, and (iii) non-redundancy with other retained variables. Specifically, maternal age was retained because adolescent and advanced maternal age are recognised obstetric risk factors; gestational age was retained to capture prematurity-related risk; gravida, parity, and number of children alive were retained to represent obstetric history and cumulative reproductive exposure; labour duration was retained because prolonged labour may contribute to uterine atony and delayed intervention; delivery method was retained because operative and caesarean delivery can alter haemorrhage risk; birth weight was retained because fetal size may influence uterine distension and delivery trauma; perineal outcome was retained to capture trauma-related bleeding risk; sex of baby was retained as a routinely recorded neonatal variable for completeness checks and exploratory modelling; previous complications and obstetric abnormalities were retained because they summarise known clinical warning signs; and the composite risk score was retained to consolidate multiple sparse binary risk indicators into one interpretable variable suitable for a small dataset.

All 13 features were then scaled to the [0, 1] range using Min-Max normalisation:

$$X\_scaled = (X - X\_min) / (X\_max - X\_min)$$

where X is the original feature value, and X_min and X_max are the minimum and maximum values of that feature in the dataset. This ensured that each feature contributed equally during training and prevented variables with larger numeric ranges from dominating the learning process; after scaling, the mean of the scaled age variable was approximately 0.45 and the mean of the scaled risk score was approximately 0.52.

2.2.3 Model Development

The target variable was a binary PPH outcome (estimated blood loss ≥ 1,000 mL), with 118 no-PPH cases and 105 PPH cases (47.09% prevalence) - a relatively balanced distribution. Using train_test_split with stratification on the target variable, the data were divided into training (178 records, 80%) and test (45 records, 20%) sets, preserving class proportions in both subsets. Features and labels were subsequently converted to PyTorch tensors for the MLP model.

2.2.3.1 Random Forest Model

A Random Forest classifier with 500 decision trees was trained, with a maximum depth of 6, a minimum of 5 samples required to split an internal node, and a minimum of 2 samples per leaf. The random_state was fixed at 42 for reproducibility and n_jobs set to -1 to parallelise training.

The class_weight parameter was set to 'balanced_subsample' rather than a fixed 'balanced' weighting. Because each tree in the ensemble is trained on a bootstrap resample of the data, the class ratio within any individual bootstrap sample can deviate from the overall dataset ratio even when the full dataset is close to balanced. 'balanced_subsample' recomputes class weights

independently for each bootstrap sample rather than applying a single fixed weighting across the whole dataset, making the ensemble more robust to this resampling variance and reducing the risk of the minority (PPH) class being under-represented in any individual tree.

2.2.3.2 MLP Model

A Multilayer Perceptron was implemented in PyTorch as a deep-learning comparison model, consisting of an input layer sized to the 13 selected features, two hidden layers of 32 and 16 neurons respectively (each with ReLU activation and dropout of 0.3), and a single-neuron output layer producing a binary logit for PPH classification.

The 32-16 two-hidden-layer configuration was selected following preliminary pilot experiments that compared single- and two-hidden-layer architectures with widths of 16, 32, and 64 neurons, evaluated on held-out validation accuracy and AUC. Wider or deeper configurations showed no consistent improvement and tended to overfit given the small sample size (n = 223), whereas the 32-16 configuration provided the best balance between model capacity and generalisation. The dropout rate of 0.3 was likewise chosen after testing values between 0.2 and 0.5: lower rates provided insufficient regularisation, while higher rates degraded validation performance by under fitting the already limited training data. An L2 weight-decay term of 1e-4 was added to the Adam optimiser as a further regularisation measure to reduce overfitting risk on the small dataset.

The model was trained with the Adam optimiser (learning rate 0.0005, weight decay 1e-4) using BCEWithLogitsLoss with a positive-class weight of 1.1 to give modest additional emphasis to the minority class, for 80 epochs with a batch size of 16. Gradient clipping (maximum norm 1.0) was applied to stabilise training, which was performed on a CUDA-enabled GPU where available, or otherwise on CPU. At inference, a sigmoid activation was applied to the output logits with a 0.5 classification threshold.

2.2.4 Model Evaluation

Both models were evaluated on the same 20% held-out test set (45 records) using accuracy, precision, recall, and F1-score:

$$Precision = TP / (TP + FP) \qquad Recall = TP / (TP + FN)$$

$$F1 = 2 \times (Precision \times Recall) / (Precision + Recall)$$

where TP, FP, and FN denote true positives, false positives, and false negatives respectively. ROC-AUC was computed to assess discriminatory power independent of classification threshold, and confusion matrices were generated to characterise the distribution of true and false positives and negatives for each model.

2.2.5 Model Interpretability

Model interpretability was assessed using SHapley Additive exPlanations (SHAP) with DeepExplainer applied to the MLP model. SHAP values were computed for the first 50 test samples against a background of 100 sampled training records. Global summary plots characterised the overall influence of each feature on model output, while individual force plots identified the features most responsible for increasing or decreasing predicted PPH risk for specific patients, with the three highest-magnitude SHAP contributors reported for each of the first five test samples. This approach added clinical transparency to the otherwise opaque neural network predictions.

2.2.6 Model Saving and Deployment Considerations

The trained MLP weights were saved as a PyTorch state dictionary (pph_prediction_model.pth), and the associated preprocessing components - the feature scaler, label encoders, and the final list of selected feature columns - were serialised with Joblib (preprocessing_objects.pkl) to ensure that any future inference pipeline applies identical transformations to new data.

Beyond storage, translating this framework into clinical use requires deliberate attention to deployment within the existing clinical workflow. A practical workflow would present the model through a lightweight interface at the point of care for example, a web or tablet based form used by midwives at admission and during labour to enter the 13 required variables, most of which are already captured in routine intrapartum records. The system would return a real-time PPH risk score together with its principal SHAP-identified contributing factors, displayed as a colour-coded (low/medium/high) alert on a ward dashboard rather than a raw probability, so that the output maps directly onto existing clinical escalation pathways. Where an electronic health record (EHR) system is in use, integration via interoperable standards such as HL7 FHIR would allow the model to draw required variables automatically and write risk scores back into the patient record, reducing duplicate data entry. Given the small size and single-site origin of the training data, any deployment should begin with a silent-mode pilot - generating predictions alongside, but not visible to, routine clinical decisions - to evaluate calibration and alert thresholds locally before active clinical alerting is enabled, together with a clear protocol for clinician override. The model's modest computational footprint (a shallow MLP and a bounded Random Forest) makes it feasible to run on low-power hardware or offline in settings with limited connectivity, which is an important consideration for the rural and peripheral facilities where PPH-related mortality is often highest.

3. Results and Discussion

3.1 Model Performance

Both models were evaluated on the 45-record test set (24 no-PPH, 21 PPH cases). The Random Forest model achieved an overall accuracy of 0.8667, with a precision, recall, and F1-score of 0.8462, 0.9167, and 0.8800 respectively for the no-PPH class, and 0.8947, 0.8095, and 0.8500 for the PPH class (macro-average F1 = 0.8650; weighted-average F1 = 0.8660), together with a ROC-AUC of 0.8730. Its confusion matrix showed 22 true negatives, 17 true positives, 2 false positives, and 4 false negatives.

The MLP model achieved a lower overall accuracy of 0.8222, with precision, recall, and F1-score of 0.8077, 0.8750, and 0.8400 for the no-PPH class, and 0.8421, 0.7619, and 0.8000 for the PPH class (macro-average F1 = 0.8200; weighted-average F1 = 0.8213), together with a ROC-AUC of 0.8115. Its confusion matrix showed 21 true negatives, 16 true positives, 3 false positives, and 5 false negatives - one additional false positive and one additional false negative relative to Random Forest.

3.2 Random Forest Feature Importance

Feature importance analysis of the Random Forest model showed Duration_of_Labour_minutes as the strongest predictor (importance ≈ 0.33), followed by Delivery_Method_Encoded (≈ 0.30); together these two variables accounted for over 60% of total importance. Age (≈ 0.09) and Perineum_Encoded (≈ 0.08) followed, with Birth_Weight_numeric (≈ 0.06), Risk_Score and Gravida (≈ 0.05 each), Children_Alive and Parity (≈ 0.03 each), and Has_Previous_C_S and Sex_of_Baby_Encoded (≈ 0.02 each) contributing more modestly. Has_Obst_Complications (≈ 0.01) and Has_Abnormalities (≈ 0.00) contributed negligibly, indicating that intrapartum and procedural variables were considerably stronger predictors in this dataset than historical complication indicators (Figure 1).

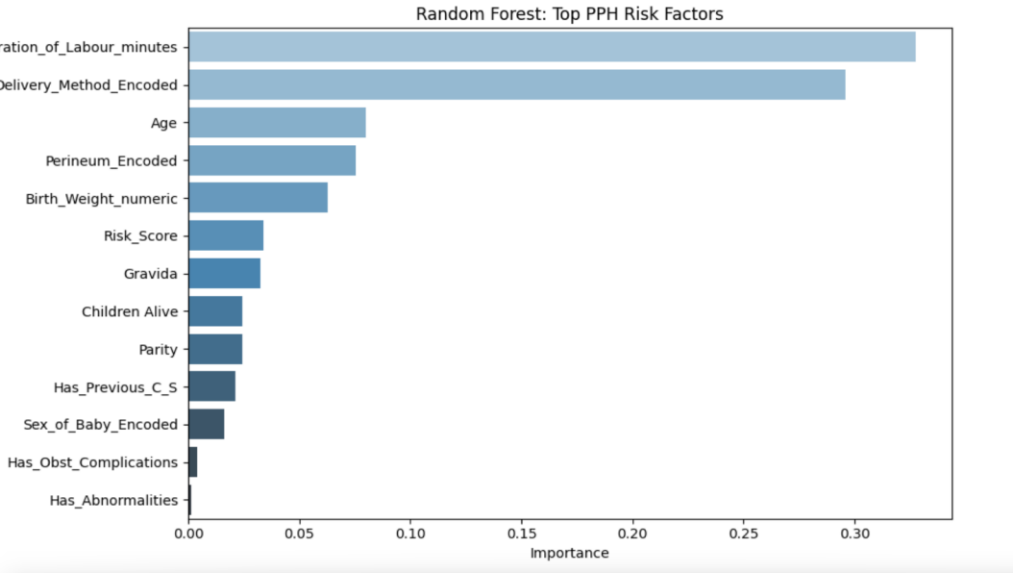

Figure 1: Random Forest feature importance ranking for PPH risk factors.

Table 1: Model comparison classification performance

Metric	RF – No PPH	RF – PPH	RF – Overall	MLP – No PPH	MLP – PPH	MLP – Overall
Precision	0.8462	0.8947	–	0.8077	0.8421	–
Recall	0.9167	0.8095	–	0.8750	0.7619	–
F1-score	0.8800	0.8500	–	0.8400	0.8000	–
Support	24	21	45	24	21	45
Accuracy	–	–	0.8667	–	–	0.8222
Macro-avg precision	–	–	0.8704	–	–	0.8249
Macro-avg recall	–	–	0.8631	–	–	0.8185

Metric	RF – No PPH	RF – PPH	RF – Overall	MLP – No PPH	MLP – PPH	MLP – Overall
Macro-avg F1	–	–	0.8650	–	–	0.8200
Weighted-avg precision	–	–	0.8688	–	–	0.8238
Weighted-avg recall	–	–	0.8667	–	–	0.8222
Weighted-avg F1	–	–	0.8660	–	–	0.8213
ROC-AUC	–	–	0.8730	–	–	0.8115

3.3 Model Comparison

Table 1 summarizes the complete classification performance of both models on the shared test set.

Random Forest outperformed the MLP across every evaluation metric: accuracy (0.8667 vs 0.8222), macro-average F1 (0.8650 vs 0.8200), weighted-average F1 (0.8660 vs 0.8213), and ROC-AUC (0.8730 vs 0.8115, a margin of 0.0615). This advantage held for both classes individually - Random Forest achieved higher precision, recall, and F1-score than the MLP for both the no-PPH and PPH classes.

Table 2: Confusion matrix comparison

Model	True Negatives	False Positives	False Negatives	True Positives	Total
Random Forest	22	2	4	17	45
MLP	21	3	5	16	45

Random Forest produced one fewer false positive and one fewer false negative than the MLP (Table 2). Since a false negative in this context represents a missed PPH case with potential delay in intervention, the lower false-negative count for Random Forest is particularly relevant to its clinical reliability. Figure 2 presents the corresponding ROC curves, confirming the higher discriminative power of the Random Forest model across the full range of classification thresholds.

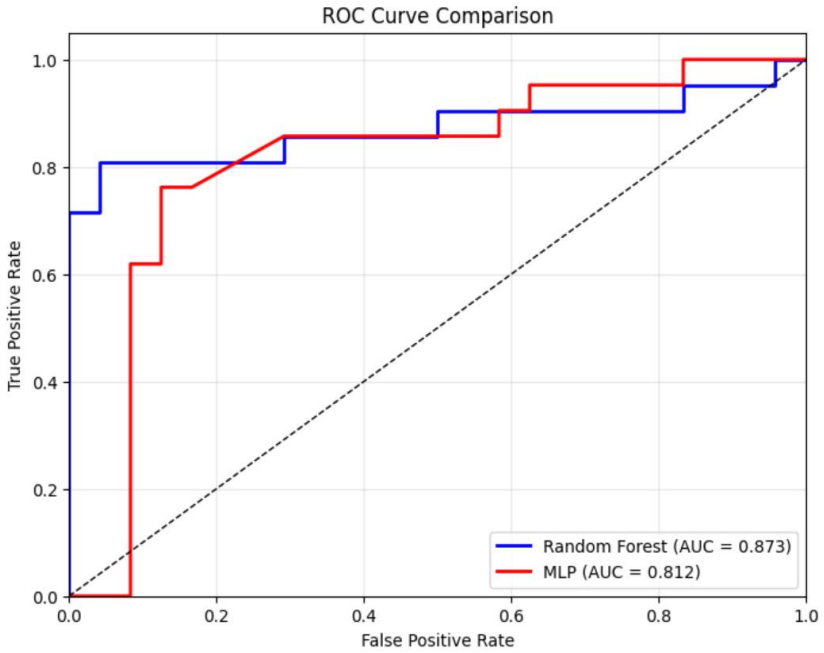

Figure 2: ROC curve comparison between Random Forest and MLP models.

Overall, Random Forest demonstrated greater discriminative power, class specificity, precision, and recall, with fewer misclassifications, and is accordingly identified as the better-performing model in this study.

consideration given that a missed PPH case can delay life-saving intervention.

Relative to the literature reviewed in the Introduction, this performance is competitive despite the comparatively small sample size. Raman et al. (2025) reported an AUC of 0.96 using larger datasets with dynamic labour parameters; Yao et al. (2025) achieved strong performance with XGBoost on 1,225 records; Lérias Cambeiro et al. (2025) obtained an AUC of 0.907 with ridge logistic regression; and Xiao et al. (2025) achieved an AUC of 0.868 by combining imaging with clinical variables. Although the AUC obtained here (0.8730) does not exceed the highest reported values, it compares favourably given that this study used only 223 records of routinely collected clinical variables, without imaging or multi centre data, suggesting that reliable predictive performance is achievable even in low-resource settings using data that are already collected as part of standard care.

A key strength of this study is its incorporation of SHAP based interpretability, which is frequently absent from accuracy-oriented PPH prediction models. Rather than reporting predictive metrics alone, the framework identifies labour duration and delivery method as the dominant, clinically plausible predictors, strengthening confidence in the model's outputs among clinical end-users. Taken together, the results support the feasibility of AI-based early warning systems in resource-limited obstetric settings, while indicating that larger, multi-centre validation is an important next step before clinical adoption.

3.5 Study Limitations

This study has several limitations that should be considered when interpreting its findings. First, and most importantly, the sample size (n = 223) is small relative to the datasets typically used in comparable ML studies (e.g., 1,225 records in Yao et al., 2025), which constrains statistical power, widens the plausible uncertainty around the reported performance estimates, and increases the risk that the models - despite regularization via dropout, weight decay, and constrained tree depth have partly fit idiosyncrasies of this particular sample rather than generalizable clinical relationships. Second, the data originate from a single tertiary hospital in Zimbabwe; case-mix, staffing, and documentation practices at other facilities, particularly rural or primary-care settings where PPH mortality is often highest, may differ substantially, limiting the generalizability of the model in its current form. Third, the exclusion of HIV status, discussed in Section 2.2.2, removes a variable of known clinical relevance in the African obstetric context and may omit some genuine predictive signal.

To address these limitations, future work should prioritize: (i) external validation on independent datasets from other low-resource maternity units, ideally spanning multiple countries and facility types, to assess generalizability; (ii) prospective, multi-site data collection to increase sample size and enable subgroup analyses (e.g., by HIV status once more completely documented data are available); (iii) reporting of bootstrap or cross-validated confidence intervals around performance metrics rather than single point estimates, to better characterize uncertainty given the current sample size; and (iv) exploration of privacy-preserving data-pooling approaches, such as federated learning across participating hospitals, to grow the effective training sample without centralizing sensitive patient data. These steps are necessary precursors to any clinical deployment of the type outlined in Section 2.2.6.

4. Conclusion

This study designed, implemented, and evaluated an AI-based early warning system for postpartum haemorrhage using routine obstetric data from Mpilo Central Hospital, Zimbabwe. Through systematic preprocessing, feature engineering, and comparison of Random Forest and MLP models, the Random Forest classifier achieved the strongest performance, with an accuracy of 0.8667, balanced precision and recall for the PPH class, and a ROC-AUC of 0.8730 on unseen test data. Intrapartum factors - particularly labour duration and delivery method - emerged as the dominant predictors of PPH risk.

These findings demonstrate that machine learning can generate clinically valuable predictions from the kind of small, routinely collected datasets common in low- and middle-income settings, with Random Forest outperforming the MLP and offering interpretable explanations via SHAP. While the resulting AUC does not match the highest values reported in larger or multimodal studies, it is competitive and contextually applicable, supporting early risk detection and proactive management strategies such as uterotonic administration, fluid replacement, or surgical pre-preparation.

By addressing gaps in interpretability, deployment readiness, and applicability to resource-constrained settings, this study contributes a practical foundation for AI assisted obstetric care. The saved model and preprocessing pipeline provide a basis for future integration into clinical decision-support systems, with the potential to reduce avoidable maternal morbidity and mortality associated with PPH in similar settings.

5. Recommendations

1. Conduct external validation of the Random Forest model on independent datasets from other low-resource maternity units to assess generalizability across different populations.
2. Expand the feature set in future iterations to include additional real-time intrapartum variables (e.g., vital-sign trends, uterine tone measurements) and psychosocial variables (e.g., fear of childbirth) that may improve predictive quality.
3. Evaluate state-of-the-art ensemble algorithms, such as XGBoost, or hybrid statistical-machine-learning models, to compare and potentially improve accuracy and calibration on similarly small datasets.
4. Develop and pilot a clinical decision-support implementation of the model within maternity wards, following the deployment pathway outlined in Section 2.2.6, to assess its effect on timely intervention and maternal outcomes.
5. Incorporate complementary interpretability methods beyond SHAP, such as LIME or counterfactual explanations, to strengthen clinician understanding of, and trust in, model predictions.
6. Investigate adaptation of the framework for mobile or low-compute environments to support use in rural or remote healthcare facilities with limited technological infrastructure.
7. Pursue multi centre partnerships to build larger, harmonized African obstetric datasets, enabling more robust model training and helping to overcome the data-scarcity constraints identified in this study.

References

Elser, H., Mumford, S., Grantz, K., Pollack, A., Mendola, P., Mills, J., Yeung, E., Zhang, C., Schisterman, E., & Hinkle, S. (2025). Postpartum Haemorrhage and Long-Term Mortality. Paediatric and Perinatal Epidemiology. https://doi.org/10.1111/ppe.13166

Lee, S., & Kang, Y. (2022). Postpartum hemorrhage. Taiwanese Journal of Obstetrics & Gynecology, 61(1), 5-7. https://doi.org/10.1016/j.tjog.2021.11.003

Riches, J., Jafali, J., Twabi, H., Chimwaza, Y., Onrust, M., Bilesi, R., Gadama, L., Kachale, F., Kuyere, A., Makhaza, L., Makuluni, R., Munthali, L., Musopole, O., Ndamala, C., Phiri, D., Coomarasamy, A., Merriel, A., Waitt, C., Odland, M., & Lissauer, D. (2025). Avoidable factors associated with maternal death from postpartum haemorrhage: a national Malawian surveillance study. BMJ Global Health, 10. https://doi.org/10.1136/bmjgh-2024-015781

Sumikura, H. (2025). Postpartum Hemorrhage: Preventable Maternal Deaths and the Role of Coagulation Monitoring. Journal of Obstetrics and Gynaecology Research, 51. https://doi.org/10.1111/jog.70134

Ende, H., & Butwick, A. (2021). Current State and Future Direction of Postpartum Hemorrhage Risk Assessment. Obstetrics & Gynecology, 138, 924-930. https://doi.org/10.1097/aog.0000000000004579

Joshi, D. (2025). Why Traditional Statistical Methods Need to Evolve in the Age of Artificial Intelligence: A Biostatistical Perspective. Nepal Journal of Public Health. https://doi.org/10.70280/njph(2025)v2i1.31

Raman, R., Kumar, V., Pillai, B., Verma, A., Rastogi, S., & Pandey, D. (2025). Predictive Machine Learning Models for Postpartum Hemorrhage Risk Assessment. 2025 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE), 1-6. https://doi.org/10.1109/iitcee64140.2025.10915212

Yao, X., Bao, Y., Wu, N., Shan, S., Xu, Y., Huo, K., Huang, R., & Ying, H. (2025). Interpretable machine-learning-based prediction of postpartum haemorrhage in normal vaginal births in Shanghai, China. Frontiers in Medicine, 12. https://doi.org/10.3389/fmed.2025.1670987

Xiao, Y., Hu, Z., Yuan, B., & Sun, Z. (2025). Postpartum Hemorrhage Prediction Based on Convolutional Neural Networks. 2025 2nd International Conference on Digital Image Processing and Computer Applications (DIPCA), 182-186. https://doi.org/10.1109/dipca65051.2025.11042290

Lérias-Cambeiro, M., Mugeiro-Silva, R., Rodrigues, A., Dias-Domingues, T., Lança, F., & Carneiro, A. (2025). Enhancing Postpartum Haemorrhage Prediction Through the Integration of Classical Logistic Regression and Machine Learning Algorithms. Mathematics. https://doi.org/10.3390/math13213376

Nelli, F. (2023). Python Data Analytics: With Pandas, NumPy, and Matplotlib. Python Data Analytics. https://doi.org/10.1007/978-1-4842-9532-8

Adugna, T., Xu, W., & Fan, J. (2022). Comparison of Random Forest and Support Vector Machine Classifiers for Regional Land Cover Mapping Using Coarse Resolution FY-3C Images. Remote Sensing, 14, 574. https://doi.org/10.3390/rs14030574

Nohara, Y., Matsumoto, K., Soejima, H., & Nakashima, N. (2021). Explanation of Machine Learning Models Using Shapley Additive Explanation and Application for Real Data in Hospital. Computer Methods and Programs in Biomedicine, 214, 106584. https://doi.org/10.1016/j.cmpb.2021.106584

Parida, S., Gerostathopoulos, I., & Bogner, J. (2025). How Do Model Export Formats Impact the Development of ML-Enabled Systems? A Case Study on Model Integration. 2025 IEEE/ACM 4th International Conference on AI Engineering - Software Engineering for AI (CAIN), 48-59. https://doi.org/10.1109/cain66642.2025.00014