Journal of Advanced Health Informatics Research (JAHIR)
Vol. 4, No. 1, (2026), pp. 48-59
ISSN: 2985-6124, DOI: 10.59247/jahir.v4i1.412

EMR-Based Multiclass Classification of Primary Childbirth Complications in High-Risk Pregnant Women Using Explainable Machine Learning

ARTICLE INFO

Article history:

Received: May 25, 2026
Revised: June 20, 2026
Published: July 31, 2026

Keywords:

Machine Learning;
Labor Complications;
SHAP;
Medical Prediction;
Optimization of Midwifery Care

ABSTRACT

Birth complications remain a major contributor to maternal morbidity and mortality in Indonesia, particularly in referral healthcare facilities where rapid clinical recognition is required. This single-center retrospective study developed an explainable Electronic Medical Record (EMR)-based multiclass classification model to classify the primary childbirth complication category among women with recorded labor complications at a tertiary referral hospital. The model was not designed to screen uncomplicated pregnancies for future complication occurrence. When more than one diagnosis was recorded, the primary complication documented in the EMR was used as the single target label, while secondary diagnoses were not modeled. The study included 10,644 maternal records from 2020 to 2024 at Margono Soekarjo General Hospital. Predictors were restricted to maternal, obstetric, clinical, and laboratory variables available at admission or pre-delivery evaluation before definitive complication classification; variables potentially occurring after clinical decision-making, such as mode of delivery, were excluded from model input. The XGBoost algorithm was used with stratified train-test splitting, SMOTE applied only to the training set, and SHAP-based interpretation. The model achieved 70% accuracy, a macro F1-score of 0.65, a weighted F1-score of 0.70, and a macro-AUC of 0.92. These results indicate promising but moderate preliminary performance, with weaker classification for hemorrhage and infection. SHAP suggested that the model relied primarily on metabolic, hemodynamic, inflammatory, and hematological indicators. External validation, temporal validation, calibration assessment, and prospective workflow evaluation are required before clinical implementation.

Corresponding Author:
Surtiningsih, Midwifery Study Program, Faculty of Health, Universitas Harapan Bangsa, Purwokerto, Central Java, Indonesia
Email: surtiningsih@uhb.ac.id

1. INTRODUCTION

The Maternal Mortality Ratio (MMR) remains a key indicator of maternal healthcare quality and continues to pose a major challenge in Indonesia despite a decline over the past few decades. Indonesia's Maternal Mortality Ratio (MMR) decreased from 450 to 249 per 100,000 live births between 1990 and 2020; this figure, however, remains above the 2030 Sustainable Development Goals (SDGs) target of 70 per 100,000 live births [1]. Regional disparities and the high burden of obstetric complications indicate that strengthening the maternal healthcare system remains critically necessary [2],[3]. Hypertensive disorders of pregnancy, preeclampsia, postpartum hemorrhage, and infections remain the leading causes of maternal morbidity and mortality, particularly in tertiary referral healthcare facilities [4],[5]. Most of these complications develop progressively and require rapid early detection so that clinical interventions can be performed in a timely manner to prevent more severe conditions [6].

The risk of childbirth complications is multifactorial and influenced by a combination of maternal, obstetric, physiological, and laboratory factors. Characteristics such as extreme maternal age, high parity, a history of abortion, multiple pregnancies, blood pressure, nutritional status, and hematological conditions play significant roles in increasing the risk of maternal and neonatal morbidity [7],[8]. In the context of referral healthcare services, Margono Soekarjo General Hospital, a tertiary referral hospital with a high volume of high-risk deliveries, possesses a comprehensive and integrated Electronic Medical Record (EMR) database. The availability of these routine clinical data provides a valuable opportunity to develop data-driven prediction systems that support the more systematic early detection of childbirth complications [3],[9].

Machine learning has increasingly been used to support obstetric risk modeling using routinely collected clinical data. Previous studies have reported models for autoimmune pregnancy complications, hypertensive disorders, birth asphyxia, postpartum hemorrhage, and preterm birth [10]-[15]. These studies demonstrate the potential of data-driven approaches for maternal care, although many have focused on a single outcome or have been developed in settings that differ from Indonesian referral hospitals. The clinical task and deployment context should therefore be clearly defined before presenting an EMR-based model as a decision-support tool.

The novelty of this study is reflected in four aspects. It uses a large routine EMR dataset from an Indonesian tertiary referral hospital, reflecting the real-world case mix of high-risk maternal care in a resource-constrained referral setting. The study frames the problem as a multiclass classification task for the primary childbirth complication category across six clinically important outcomes rather than as separate single-outcome models. The model integrates SHAP-based interpretation to examine how the trained classifier uses routinely available maternal, obstetric, clinical, and laboratory variables. The study also clarifies the potential EMR integration point as a preliminary classification aid for clinicians reviewing high-risk or symptomatic parturient women rather than as an autonomous diagnostic system.

XGBoost was selected because it is well suited for structured tabular EMR data, can model nonlinear interactions among predictors, and includes regularization mechanisms that may reduce overfitting [18], [19]. Its tree-based structure is also compatible with SHAP, which enables transparent post hoc interpretation of feature contributions [20]. In this study, XGBoost was used as a pragmatic modelling approach for retrospective EMR data rather than as evidence of clinical deployment readiness.

Accordingly, this study developed an EMR-based XGBoost model to classify the primary complication category among women with recorded childbirth complications at Margono Soekarjo General Hospital. The target task was not to predict whether an uncomplicated woman would develop a complication. Instead, the model assigned one principal label from six categories: hypertension, preeclampsia, infection, premature rupture of membranes, hemorrhage, and prolonged labor. When multiple diagnoses were present in the EMR, the primary recorded diagnosis was used as the multiclass target, whereas secondary diagnoses were not modeled. This design was chosen because the available dataset contained one principal complication label per analytical case. A multi-label model would be more appropriate in future studies if complete secondary diagnosis labels become consistently available.

2. METHODS

2.1. Study Design and Location

This study employed an analytical observational approach with a retrospective cohort design based on secondary EMR data from 2020 to 2024. The study was conducted at Margono Soekarjo Regional General Hospital, Banyumas, Central Java, a Type A tertiary referral hospital with a high volume of high-risk delivery cases. The prediction point was defined as the stage at which maternal, obstetric, clinical, and laboratory information was available from admission or pre-delivery evaluation before definitive complication classification. All analyses used de-identified secondary EMR data. Ethical approval and institutional permission for secondary EMR use were obtained before analysis; the ethics approval number should be completed by the authors before final submission. Direct patient identifiers were removed before data processing, and the dataset was analyzed only in aggregated form.

2.2. Population and Sample

The initial population consisted of 10,896 delivery records with documented labor-complication diagnoses in the EMR system during the 2020 to 2024 period. Uncomplicated deliveries were not included in the analytical dataset. Therefore, the model should be interpreted as a classifier of the primary complication type among complicated cases, not as a screening model for predicting whether all high-risk pregnant women will develop a complication. Inclusion criteria were parturient women with a primary recorded diagnosis of hypertension, preeclampsia, hemorrhage, infection, premature rupture of membranes, or prolonged labor, and with relevant maternal, obstetric, clinical, and laboratory information available prior to definitive complication classification. Records without a usable primary complication label, duplicated records, incomplete essential predictors, or inconsistent clinical information were excluded. After selection, validation, and cleaning, 10,644 cases were retained as the final analytical dataset.

Data cleaning included format standardization, duplicate checking, consistency evaluation, and missing-value handling according to variable type. Observations without a primary outcome label were excluded because they could not be used in supervised multiclass learning. Numerical missing values were handled using median imputation, whereas categorical missing values were retained using a constant “Missing” category to avoid unnecessary case deletion. Target classes with extremely small observation counts were removed before splitting to preserve stratified sampling validity. The analytical flow from the original EMR extraction to the final modelling dataset is summarized below in Table 1.

Table 1. Data Cleaning Flow and Analytical Dataset Derivation

Stage	Description	Number of Records
Initial EMR extraction	Delivery records with documented labor-complication diagnoses during 2020 to 2024	10.896
Excluded during selection, validation, and cleaning	Records removed because of missing primary label, duplicated entries, incomplete essential predictors, or inconsistent clinical information	252
Final analytical dataset	Records retained for multiclass modelling	10.644
Training set before SMOTE	Stratified 80% split used for model training and resampling	8.515
Held-out test set	Stratified 20% split preserved without SMOTE	2.129
Training set after SMOTE	Oversampled training set used only for model fitting	13.836

2.3. Research Variables

Predictor variables were derived from routine EMR data and grouped into maternal, obstetric, clinical, and laboratory domains. Maternal variables included age, education, and occupation. Obstetric variables included gravidity, parity, history of abortion, gestational age, fetal presentation, and pregnancy type. Clinical variables consisted of systolic blood pressure, diastolic blood pressure, and body mass index (BMI). Laboratory variables included hemoglobin level, white blood cell count, and platelet count. Variables were selected only if they were available at admission or during the pre-delivery evaluation before definitive complication classification. Mode of delivery was not used as a model predictor because it may occur after clinical decision-making and could introduce temporal leakage.

Table 2. Predictor Availability, Timing of Measurement, and Leakage Control

Domain	Variables	Timing of Measurement	Leakage Control
Maternal	Age, education, occupation	Registration or admission record	Available before complication classification
Obstetric	Gravidity, parity, abortion history, gestational age, fetal presentation, type of pregnancy	Antenatal history or admission/pre-delivery assessment	Included only when documented before definitive complication label
Clinical	Systolic blood pressure, diastolic blood pressure, BMI	Admission or pre-delivery clinical assessment	Used as pre-classification clinical information
Laboratory	Hemoglobin, white blood cell count, platelet count	Initial laboratory examination during admission or pre-delivery evaluation	Used only when available before definitive complication classification
Derived features	MAP, PP, anemia, leukocytosis, thrombocytopenia, leukocyte-to-platelet ratio, hemoglobin-to-platelet ratio	Calculated from the variables above	Derived only from pre-classification predictors
Excluded from model input	Mode of delivery and other post-decision variables	After clinical decision-making or delivery process	Excluded to reduce temporal leakage

Feature engineering was performed through text cleaning, format standardization, categorical transformation, and the creation of clinically derived features. Occupation was simplified into major categories, education was transformed into an ordinal scale, and gestational age was converted into preterm, term, and postterm categories. BMI was classified into nutritional status categories, whereas laboratory parameters were transformed into indicators such as anemia, leukocytosis, and thrombocytopenia. Mean arterial pressure (MAP), pulse pressure (PP), the leukocyte-to-platelet ratio, and the hemoglobin-to-platelet ratio were calculated from variables available before the prediction point. These derived features were used to strengthen the representation of maternal hemodynamic, inflammatory, and hematological profiles without using post-diagnosis variables.

2.4. Research Stages

Figure 1 illustrates the study workflow, beginning with the extraction of secondary EMR data from Margono Soekarjo General Hospital for the 2020 to 2024 period. The workflow included data selection, cleaning, standardization, variable transformation, and feature engineering. The dataset was then divided into training data (80%) and held-out test data (20%) using stratified splitting to preserve the proportion of each primary complication category. The test set was kept unchanged throughout the analysis to maintain an objective estimate of model performance under realistic class distribution.

SMOTE was applied only after train-test splitting and only to the training data. This procedure was used to improve representation of minority classes during model fitting while preventing synthetic samples from leaking into the held-out test set. The XGBoost multiclass classifier was trained on the resampled training data and evaluated on the original test data. Model interpretation was conducted using SHAP to examine how the trained model used predictor variables. The resulting system should be interpreted as a preliminary EMR-based classification aid that requires temporal validation, external validation, calibration assessment, and prospective workflow evaluation before clinical use.

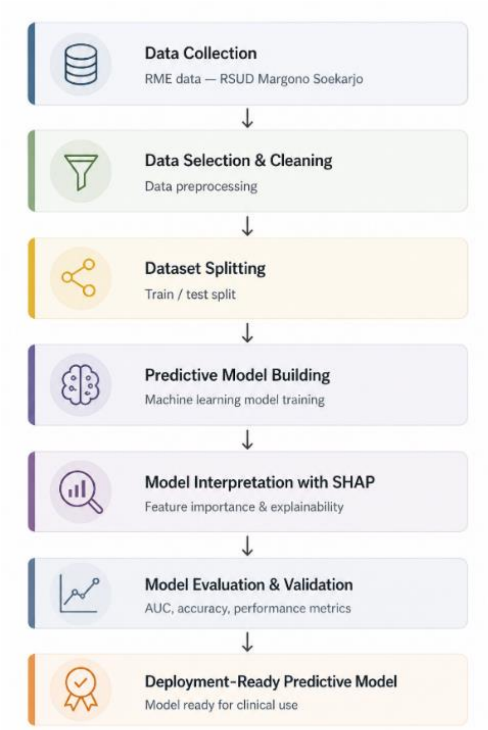

Fig. 1. Research Workflow of Multiclass Prediction Model Development

2.5. Development of the XGBoost Model

XGBoost was applied as a tree-based gradient boosting classifier for six-class prediction using structured EMR variables because of its strong predictive capability for multiclass classification and structured tabular data [22], [23]. The model was trained with a multiclass soft-probability objective so that each case received a probability score for each complication category, and the final predicted label was assigned to the class with the highest probability. Model development used the training set only, with categorical encoding, imputation, SMOTE-based resampling, and hyperparameter tuning performed within the training workflow. The tuned hyperparameters included the learning rate, number of estimators, maximum tree depth, subsampling ratio, column subsampling ratio, minimum child weight, gamma, and L1/L2 regularization. A fixed random seed was used to support reproducibility.

In this study, the multiclass label represented the primary recorded complication. The model therefore classified one principal category per case rather than estimating multiple simultaneous complications. This choice was made because the available analytical dataset provided one primary label for each record. If future EMR extraction provides complete secondary diagnosis labels, a multi-label modeling strategy should be considered.

The choice of XGBoost was based on its suitability for tabular clinical data, its capacity to model nonlinear relationships, and its compatibility with SHAP-based interpretation. The model should not be interpreted as clinically deployable solely based on retrospective performance. Additional temporal validation, external validation, and calibration are needed to determine its generalizability and clinical reliability.

2.6. Model Performance Evaluation

Model performance was evaluated on the original held-out test set using accuracy, precision, recall, F1-score, the confusion matrix, the ROC curve, and AUC, which are widely recommended for evaluating imbalanced multiclass machine learning models in medical applications [25]–[27]. Macro-average and weighted-average scores were reported to distinguish balanced class-level performance from performance influenced by majority classes, providing a more comprehensive assessment of multiclass classifier performance [29]. This distinction is important because the dataset contains imbalanced complication categories.

ROC-AUC was interpreted cautiously because AUC may remain high even when recall or the F1-score is modest in imbalanced clinical datasets. Likewise, the confusion matrix provides complementary information by revealing class-specific misclassification patterns that cannot be captured by summary performance metrics alone [28]. Class-wise precision, recall, F1-score, the confusion matrix, and the Precision-Recall curve were used as complementary evaluation metrics [21]. Calibration analysis and confidence intervals were not available in the current manuscript and should be included in future validation work using retained prediction probabilities.

The evaluation focuses on preliminary discriminatory performance rather than clinical readiness. A model intended for maternal care implementation would require prospective testing, calibration assessment, decision-curve analysis, and workflow evaluation involving obstetricians and midwives.

2.7. SHapley Additive exPlanations (SHAP)

SHAP was used as a post hoc interpretability method to explain how the trained XGBoost model used predictor variables through Shapley-based feature attribution, enabling transparent interpretation of model behavior [32]. SHAP values were computed on the held-out test set following recent studies combining XGBoost with explainable artificial intelligence for clinical prediction [33]. Because the model was multiclass, SHAP explanations were interpreted in a class-specific manner for each complication category, and global importance was summarized using the mean absolute SHAP value across test-set observations. The waterfall plot was interpreted as an individual, class-specific explanation for the predicted category. SHAP values describe model behavior and feature contribution patterns; they should not be interpreted as evidence of causal clinical effects.

3. RESULTS AND DISCUSSION

3.1. Dataset Description

The final analytical dataset consisted of 10,644 cases of parturient women with one primary recorded labor-complication category: hypertension, infection, premature rupture of membranes, hemorrhage, prolonged labor, or preeclampsia. Because uncomplicated deliveries were not included, the results should be interpreted as classification of primary complication categories among documented complicated cases. The data distribution varied across classes, indicating class imbalance in the original dataset and reflecting the uneven prevalence of complication categories in referral-hospital EMR data.

After data selection, standardization, transformation, and feature engineering, the dataset produced 41 final features representing maternal, obstetric, clinical, laboratory, and derived physiological indicators. SMOTE was applied only to the training set, producing 13,836 observations in the resampled training data, while the 2,129 test cases retained their original distribution. This approach allowed the model to learn from a more balanced training set while preserving realistic evaluation conditions in the held-out test set.

3.2. Performance of the XGBoost Multiclass Model

The XGBoost multiclass model achieved an overall accuracy of 70%, with a macro F1-score of 0.65 and a weighted F1-score of 0.70. These values indicate preliminary and moderate classification performance rather than clinical readiness. The difference between the macro-average and weighted-average results indicates that the model performed better for classes with stronger representation, whereas minority or clinically heterogeneous classes remained more difficult to classify. Model interpretation should not rely on accuracy alone.

Table 3. Results of the Performance Evaluation of the Model for Classifying Primary Childbirth Complication Categories

Type of Complication	Precision	Recall	F1-Score	Number of Data Points (n)
Hypertension	0.63	0.75	0.69	494
Infections	0.49	0.62	0.55	86
Premature Rupture of Membranes	0.71	0.68	0.70	285
Bleeding	0.50	0.47	0.49	123
Prolonged Labor	0.75	0.70	0.72	576
Preeclampsia	0.80	0.72	0.75	565
Model Accuracy	-	-	0.70	2,129
Macro Average	0.65	0.66	0.65	2,129
Weighted Average	0.70	0.70	0.70	2,129

The best performance was observed for preeclampsia, with a precision of 0.80, a recall of 0.72, and an F1-score of 0.75, followed by prolonged labor with an F1-score of 0.72 and premature rupture of membranes with an F1-score of 0.70. These categories may have more consistent clinical patterns or stronger representation in the dataset. Performance was lower for infection (F1-score 0.55) and hemorrhage (F1-score 0.49). These lower scores are clinically important because both complications may progress rapidly and require timely intervention. The limited performance in these categories indicates that the current model should not be considered a reliable standalone clinical support tool for these outcomes.

A recall of 0.62 for infection indicates moderate sensitivity, whereas a precision of 0.49 suggests a substantial number of false-positive predictions. A hemorrhage F1-score of 0.49 and a recall of 0.47 indicate limited sensitivity and unstable classification. These findings may reflect smaller class sizes, the heterogeneity of clinical manifestations, overlap with other systemic conditions, and the limited ability of routine variables to capture acute bleeding dynamics. Future modeling should consider additional predictors, improved phenotyping, temporal features, and class-specific threshold optimization.

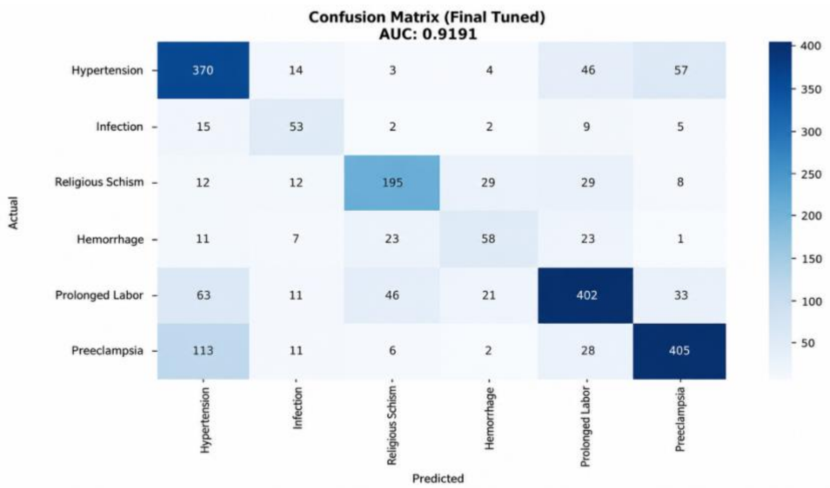

Fig. 2. Confusion Matrix Heatmap of Childbirth Complication Prediction Model

The confusion matrix shown in Figure 2 illustrates overlapping prediction patterns across the complication categories. Hypertension and preeclampsia may be misclassified because both are influenced by related hemodynamic profiles, whereas infection and hemorrhage may overlap with systemic physiological responses that are not fully differentiated by routine laboratory variables alone. These results support the feasibility of EMR-based classification but also demonstrate that additional validation and model refinement are necessary before clinical implementation.

3.3. ROC Curve and AUC Analysis

ROC curves and AUC were used to evaluate discriminatory ability across complication categories and to assess the ranking performance of the classifier across different decision thresholds [30], [31]. The ROC curves in Figure 3 were above the diagonal reference line, indicating that the model demonstrated useful discrimination across varying thresholds. AUC was not interpreted as sufficient evidence of clinical reliability because imbalanced multiclass medical data may produce high AUC values despite modest F1-scores, recall, or precision for minority classes.

The AUC values for each complication category ranged from 0.90 to 0.94, with a micro-average AUC of 0.93 and a macro-average AUC of 0.92. These values suggest strong ranking ability but should be interpreted alongside the more moderate macro F1-score and the limited performance for hemorrhage and infection. The discrepancy between the AUC and F1-score indicates that the model can often rank cases correctly but may still produce suboptimal final class assignments under a fixed threshold or the argmax decision rule.

The Precision-Recall curve in Figure 3 provides additional insight into class performance under data imbalance. Class-wise PR-AUC, confidence intervals, and calibration metrics were not available in the current analysis. This limitation reduces the ability to assess probability reliability. Future work should report class-wise PR-AUC, bootstrapped confidence intervals, calibration plots, expected calibration error, and decision-curve analysis to support a more clinically meaningful evaluation.

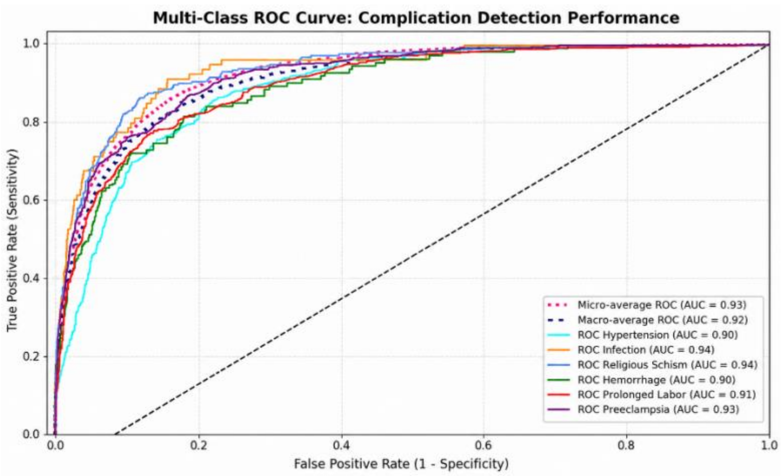

Fig. 3. ROC Curve and Precision–Recall of Childbirth Complication Prediction Model

3.4. Feature Importance Analysis

Feature importance analysis was used to evaluate the relative contribution of each feature to the XGBoost model. Based on Figure 4, the five features with the highest contribution were BMI, white blood cell count, hemoglobin-to-platelet ratio (Hb/Platelet), hemoglobin level, and pulse pressure (PP). The term BMI is used consistently throughout the manuscript to refer to Body Mass Index.

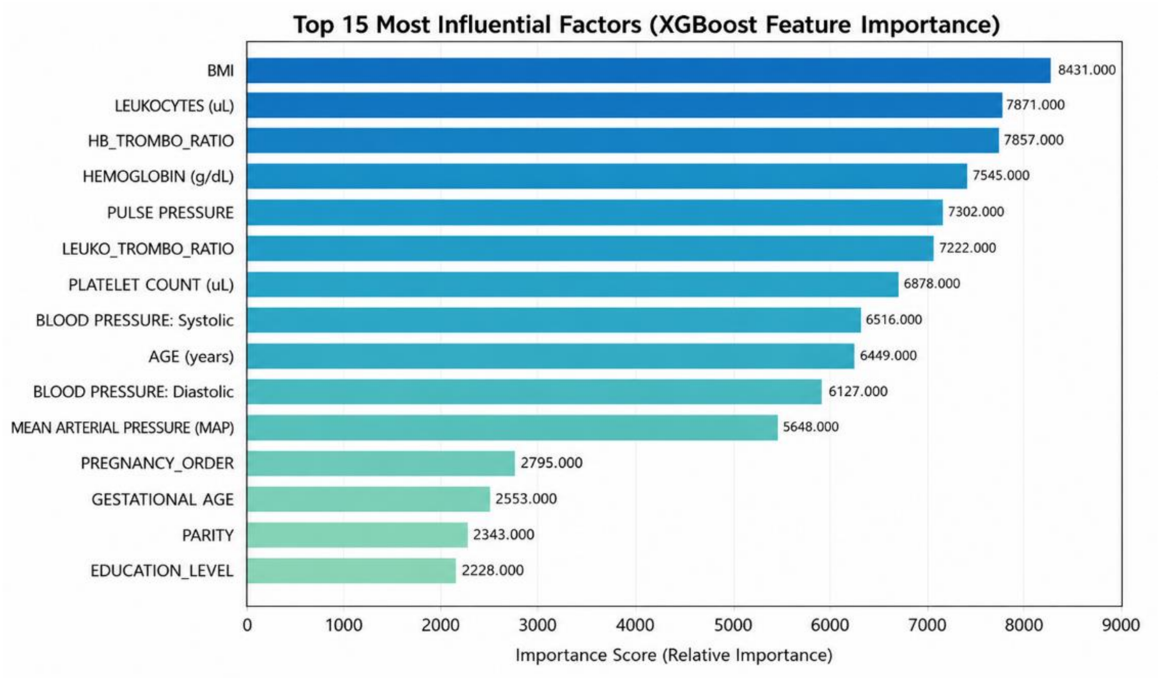

Fig. 4. XGBoost Feature Importance (Top 15 Features)

BMI emerged as the most influential model feature, suggesting that the classifier used maternal nutritional and metabolic information to differentiate among complication categories. White blood cell count contributed inflammatory information, whereas hemoglobin and the Hb/Platelet ratio contributed hematological information related to anemia, oxygen-carrying capacity, bleeding tendency, and hemostatic stability. Pulse pressure provided hemodynamic information relevant to vascular pressure patterns. These interpretations describe how the model used these variables and should not be interpreted as causal claims.

Collectively, the feature importance pattern indicates that the model used information from multiple clinical domains rather than relying on a single predictor. Feature importance reflects predictive contributions within the fitted model only. It does not establish causal relationships or confirm that modifying these variables would change the risk of complications. SHAP analysis was therefore used as an additional interpretability method to examine the direction and magnitude of feature contributions at both the global and individual levels.

3.5. SHAP-Based Model Interpretability

SHAP analysis was performed on the held-out test set to examine how predictor variables contributed to the trained XGBoost model, consistent with recent explainable AI frameworks developed for clinical prediction tasks [34]. In the multiclass setting, SHAP values were interpreted by class, and the global summary was based on mean absolute SHAP values across test-set observations. As shown in Figure 5, BMI, parity, history of abortion, white blood cell count, gravidity, hemoglobin, MAP, and systolic blood pressure contributed to the model predictions. These findings indicate that the classifier relied on multidimensional information from maternal, obstetric, hemodynamic, inflammatory, and hematological domains, consistent with previous XGBoost-SHAP studies on clinical prediction [35]. The interpretation is limited to model behavior and should not be considered proof of biological causality.

At the individual level, Figure 6 illustrates a waterfall explanation for one test-set patient predicted as hypertension. The prediction was driven mainly by systolic blood pressure, diastolic blood pressure, MAP, BMI, gestational age, and the Hb/Platelet ratio, while other variables made smaller negative contributions. This result shows how SHAP can make an individual prediction more transparent. Similar SHAP-based factor analysis has demonstrated that local feature attribution can improve the interpretability of machine learning predictions while maintaining clinician oversight during decision-making [36].

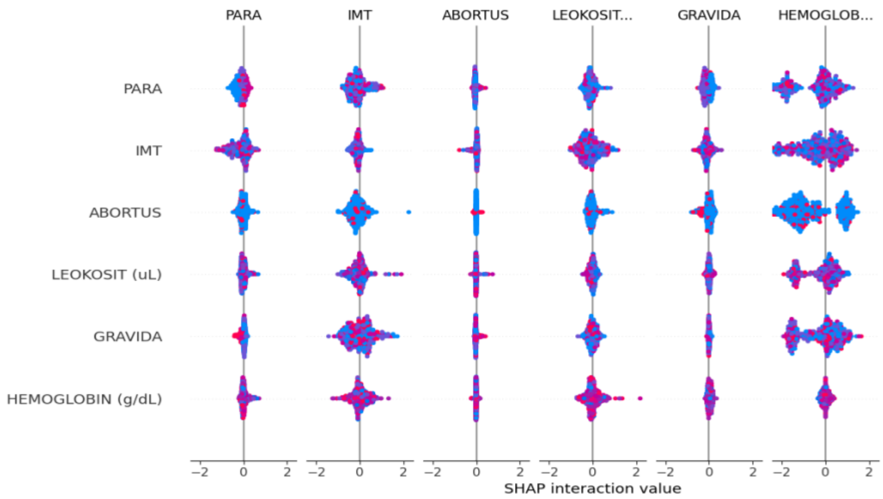

Fig. 5. SHAP Summary Plot

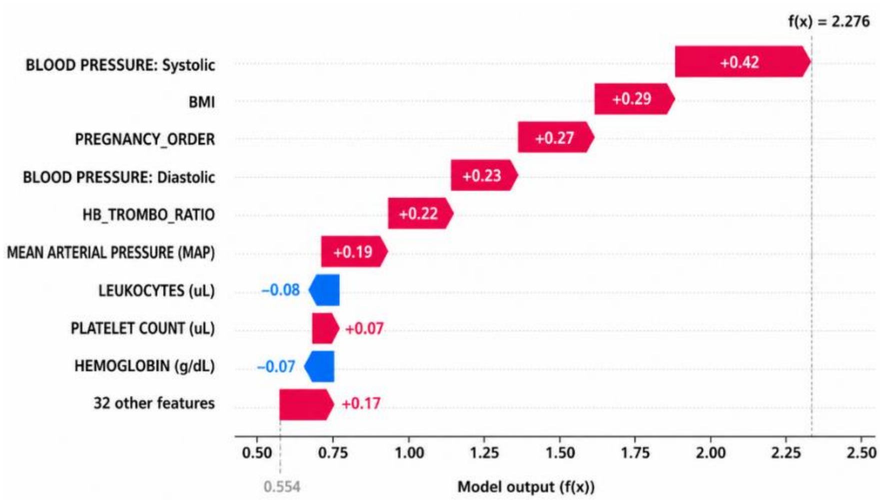

Fig. 6. SHAP Waterfall Plot of Patient 10 – Prediction: Hypertension

In a practical workflow, the model would be positioned as a non-autonomous decision-support component within the EMR. After a clinician or midwife reviews a high-risk or symptomatic parturient, the system could display the most likely primary complication category, class probabilities, and SHAP-based feature contributions. The output should be used to prompt closer clinical review, documentation verification, or referral prioritization rather than to replace obstetric assessment or determine treatment automatically. Before clinical implementation, the model requires prospective validation, calibration, usability testing, and governance procedures to ensure safe integration into clinical practice.

4. CONCLUSION

This study provides preliminary evidence that routine EMR data from a tertiary referral hospital can be used to develop an explainable multiclass model for classifying six primary childbirth complication categories among women with documented labor complications. The XGBoost model achieved an accuracy of 70%, a macro F1-score of 0.65, a weighted F1-score of 0.70, and a macro-AUC of 0.92. These findings indicate promising discriminatory performance but only moderate balanced classification performance.

The integration of XGBoost and SHAP improved the transparency of model interpretation by showing which maternal, obstetric, clinical, laboratory, and derived features contributed to the model predictions. SHAP explains model behavior rather than causal relationships. The model should therefore be viewed as an exploratory and preliminary EMR-based classification tool rather than a clinically ready decision-support system.

The main limitations of this study include its retrospective single-center design, the exclusion of uncomplicated deliveries, the use of one primary complication label per case, potential EMR recording bias, a moderate F1-score, limited performance for hemorrhage and infection, the lack of temporal and external validation, the absence of calibration assessment, and the absence of prospective workflow evaluation. Future research should include multicenter and temporal validation, multi-label modeling when secondary diagnoses are available, probability calibration, class-wise PR-AUC with confidence intervals, decision-curve analysis, and prospective evaluation in real maternal care workflows.

Acknowledgments

The authors gratefully acknowledge the financial support provided by the Beginner Lecturer Research Grant (Penelitian Dosen Pemula/PDP) Fiscal Year 2025 from the Ministry of Higher Education, Science, and Technology of the Republic of Indonesia. This research was funded under Master Contract **No. 127/C3/DT.05.00/PL/2025**, Derivative Contract **No. 044/LL6/PL/AL.04/2025**, and Institutional Contract No. **012/LPPM/PEN/06/2025**. The authors sincerely appreciate the Ministry and all supporting institutions for their assistance, which made this research possible.

AI Tools Declaration

The authors used AI-based tools (ChatGPT) for language refinement, editorial revision, and preparation of the response-to-reviewers document under author supervision. The authors remain responsible for the scientific content, clinical interpretation, data analysis, and final manuscript approval.

REFERENCES

[1] M. Syairaji, D. S. Nurdiati, B. S. Wiratama, Z. D. Prüst, K. W. M. Bloemenkamp, And K. J. C. Verschueren, "Trends And Causes Of Maternal Mortality In Indonesia: A Systematic Review," *Bmc Pregnancy Childbirth*, Vol. 24, No. 1, P. 515, Jul. 2024, Doi: 10.1186/S12884-024-06687-6.

[2] F. Fadjryani *Et Al.*, "Modeling The Duration Of Maternal Labor At Anutapura Hammer Hospital Using Lin-Ying Additive Hazard Regression," *Barekeng: Jurnal Ilmu Matematika Dan Terapan*, Vol. 20, No. 1, Pp. 0523–0540, Nov. 2025, Doi: 10.30598/barekengvol20iss1pp0523-0540.

[3] S. Martafiyah, C. Supriadi, And T. Saifudin, "Modeling The Maternal Mortality Rate In Indonesia Using Geographically Weighted Poisson Regression Approach," 2023, P. 080011. Doi: 10.1063/5.0181066.

[4] M. I. A. Akbar *Et Al.*, "The Inapres (Indonesia Preeclampsia Study): Impact Of Eclampsia On Maternal And Perinatal Outcomes Among Women With Preeclampsia: Findings From A National Multicenter Cohort," *Ajog Global Reports*, Vol. 6, No. 1, P. 100596, Feb. 2026, Doi: 10.1016/j.xagr.2025.100596.

[5] V. Pebri Giena, R. Maiseptya Sari, F. Nuraini Sasmita, And I. Rahmawati, "Developing A Mobile Application As A Learning Medium On Maternal Emergencies Related To Postpartum Hemorrhage," *Journal Of Public Health And Development*, Vol. 20, No. 1, Pp. 38–50, Jan. 2022, Doi: 10.55131/jphd/2022/200104.

[6] A. Yoseph, Y. S. Berego, M. Belayneh, And F. Guillen-Grima, "Explainable Machine Learning Algorithms For Predicting Maternal Obstetric Complications In Ethiopia: Evidence From A Retrospective Cohort In Southern Ethiopia," *Health Informatics J.*, Vol. 32, No. 1, Jan. 2026, Doi: 10.1177/14604582261428317.

[7] A. C. Gimovsky, S. S. Rasiah, C. Vergara-Lopez, P. Has, N. K. Ayala, And L. R. Stroud, "Psyche: The 5th 'P' And Its Associated Impact On The Second Stage Of Labor.," *R. I. Med. J. (2013)*, Vol. 107, No. 1, Pp. 37–44, Jan. 2024.

[8] A. Rahmawati, R. T. D. Judistiani, D. A. Setiawati, R. Fajri, S. Putri, And R. Ruslami, "Pulmonary Tuberculosis Cases In Pregnancy: Perinatal Outcomes And Health Promotion Challenges In Indonesia," *The Journal Of Infection In Developing Countries*, Vol. 16, No. 08, Pp. 1372–1375, Aug. 2022, Doi: 10.3855/Jidc.15476.

[9] L. Vasudevan *Et Al.*, "Machine Learning Models To Predict Risk Of Maternal Morbidity And Mortality From Electronic Medical Record Data: Scoping Review," *J. Med. Internet Res.*, Vol. 27, Pp. E68225–E68225, Aug. 2025, Doi: 10.2196/68225.

[10] X. Zhou *Et Al.*, "Machine Learning Techniques For Prediction In Pregnancy Complicated By Autoimmune Rheumatic Diseases: Applications And Challenges," *Int. Immunopharmacol.*, Vol. 134, P. 112238, Jun. 2024, Doi: 10.1016/J.Intimp.2024.112238.

[11] Y. Chen *Et Al.*, "Machine-Learning Predictive Model Of Pregnancy-Induced Hypertension In The First Trimester," *Hypertension Research*, Vol. 46, No. 9, Pp. 2135–2144, Sep. 2023, Doi: 10.1038/S41440-023-01298-8.

[12] F. Darsareh, A. Ranjbar, M. V. Farashah, V. Mehrnoush, M. Shekari, And M. S. Jahromi, "Application Of Machine Learning To Identify Risk Factors Of Birth Asphyxia," *Bmc Pregnancy Childbirth*, Vol. 23, No. 1, P. 156, Mar. 2023, Doi: 10.1186/S12884-023-05486-9.

[13] M. Akazawa, K. Hashimoto, N. Katsuhiko, And Y. Kaname, "Machine Learning Approach For The Prediction Of Postpartum Hemorrhage In Vaginal Birth," *Sci. Rep.*, Vol. 11, No. 1, P. 22620, Nov. 2021, Doi: 10.1038/S41598-021-02198-Y.

[14] Q. Sun *Et Al.*, "Machine Learning-Based Prediction Model Of Preterm Birth Using Electronic Health Record," *J. Healthc. Eng.*, Vol. 2022, Pp. 1–12, Apr. 2022, Doi: 10.1155/2022/9635526.

[15] R. Raja, I. Mukherjee, And B. K. Sarkar, "A Machine Learning-Based Prediction Model For Preterm Birth In Rural India," *J. Healthc. Eng.*, Vol. 2021, Pp. 1–11, Jun. 2021, Doi: 10.1155/2021/6665573.

[16] Y. Shi *Et Al.*, "Ultrasound-Based Radiomics Xgboost Model To Assess The Risk Of Central Cervical Lymph Node Metastasis In Patients With Papillary Thyroid Carcinoma: Individual Application Of Shap," *Front. Oncol.*, Vol. 12, Aug. 2022, Doi: 10.3389/Fonc.2022.897596.

[17] K. Wang *Et Al.*, "Interpretable Prediction Of 3-Year All-Cause Mortality In Patients With Heart Failure Caused By Coronary Heart Disease Based On Machine Learning And Shap," *Comput. Biol. Med.*, Vol. 137, P. 104813, Oct. 2021, Doi: 10.1016/J.Compbiomed.2021.104813.

[18] P. Zhang, Y. Jia, And Y. Shang, "Research And Application Of Xgboost In Imbalanced Data," *Int. J. Distrib. Sens. Netw.*, Vol. 18, No. 6, P. 155013292211069, Jun. 2022, Doi: 10.1177/15501329221106935.

[19] T. Chen And C. Guestrin, "Xgboost: A Scalable Tree Boosting System," In Proceedings Of The 22nd Acm Sigkdd International Conference On Knowledge Discovery And Data Mining, 2016, Pp. 785–794, Doi: 10.1145/2939672.2939785.

[20] M. Liu, C. Guo, And S. Guo, "An Explainable Knowledge Distillation Method With Xgboost For Icu Mortality Prediction," *Comput. Biol. Med.*, Vol. 152, P. 106466, Jan. 2023, Doi: 10.1016/J.Compbiomed.2022.106466.

[21] O. Rainio, J. Teuho, And R. Klén, "Evaluation Metrics And Statistical Tests For Machine Learning," *Sci. Rep.*, Vol. 14, No. 1, P. 6086, Mar. 2024, Doi: 10.1038/S41598-024-56706-X.

[22] Q. Ren And J. Wang, "Research On Enterprise Digital-Level Classification Based On Xgboost Model," *Sustainability*, Vol. 15, No. 3, P. 2699, Feb. 2023, Doi: 10.3390/Su15032699.

[23] F. Baldo, J. Grando, Y. Yamada Correa, And D. Amorim Policarpo, "Adaptive Fast Xgboost For Multiclass Classification," *Journal Of Information And Data Management*, Vol. 14, No. 1, Oct. 2023, Doi: 10.5753/Jidm.2023.3150.

[24] S. Lonang, A. Fatoni, D. Putra, F. Syuhada, A. A. Firdaus, And A. Masitha, "Hybrid Feature Selection For Effective Heart Disease Detection: A Multi-Algorithm Machine Learning Approach," *Scientific Journal Of Informatics*, Vol. 13, No. 1, 2026, Doi: 10.15294/Sji.V13i1.38815.

[25] S. A. Hicks *Et Al.*, "On Evaluation Metrics For Medical Applications Of Artificial Intelligence," *Sci. Rep.*, Vol. 12, No. 1, P. 5979, Apr. 2022, Doi: 10.1038/S41598-022-09954-8.

[26] J. S. Aguilar-Ruiz And M. Michalak, "Classification Performance Assessment For Imbalanced Multiclass Data," *Sci. Rep.*, Vol. 14, No. 1, P. 10759, May 2024, Doi: 10.1038/S41598-024-61365-Z.

[27] G. M. Foody, "Challenges In The Real World Use Of Classification Accuracy Metrics: From Recall And Precision To The Matthews Correlation Coefficient," *Plos One*, Vol. 18, No. 10, P. E0291908, Oct. 2023, Doi: 10.1371/Journal.Pone.0291908.

[28] M. Heydarian, T. E. Doyle, And R. Samavi, "Mlcm: Multi-Label Confusion Matrix," *Ieee Access*, Vol. 10, Pp. 19083–19095, 2022, Doi: 10.1109/Access.2022.3151048.

[29] K. Singh, "Product Review Classification Using Machine Learning And Statistical Data Analysis," *International Journal Of Recent Technology And Engineering (Ijrte)*, Vol. 12, No. 2, Pp. 91–96, Jul. 2023, Doi: 10.35940/Ijrte.A7530.0712223.

[30] A. M. Carrington *Et Al.*, "Deep Roc Analysis And Auc As Balanced Average Accuracy, For Improved Classifier Selection, Audit And Explanation," *Ieee Trans. Pattern Anal. Mach. Intell.*, Vol. 45, No. 1, Pp. 329–341, Jan. 2023, Doi: 10.1109/Tpami.2022.3145392.

[31] R. Numminen, I. Montoya Perez, I. Jambor, T. Pahikkala, And A. Airola, "Quicksort Leave-Pair-Out Cross-Validation For Roc Curve Analysis," *Comput. Stat.*, Vol. 38, No. 3, Pp. 1579–1595, Sep. 2023, Doi: 10.1007/S00180-022-01288-3.

[32] Y. Ning *Et Al.*, "Shapley Variable Importance Cloud For Interpretable Machine Learning," *Patterns*, Vol. 3, No. 4, P. 100452, Apr. 2022, Doi: 10.1016/J.Patter.2022.100452.

[33] A. Moore And M. Bell, "Xgboost, A Novel Explainable Ai Technique, In The Prediction Of Myocardial Infarction: A Uk Biobank Cohort Study," *Clin. Med. Insights Cardiol.*, Vol. 16, Jan. 2022, Doi: 10.1177/11795468221133611.

[34] J. Chen, J. Xiong, Y. Wang, Q. Xin, And H. Zhou, "Implementation Of An Ai-Based Mrd Evaluation And Prediction Model For Multiple Myeloma," Feb. 2024.

[35] F. Yi *Et Al.*, "Xgboost-Shap-Based Interpretable Diagnostic Framework For Alzheimer's Disease," *Bmc Med. Inform. Decis. Mak.*, Vol. 23, No. 1, P. 137, Jul. 2023, Doi: 10.1186/S12911-023-02238-9.

[36] S. U. Chowdhury, S. Sayeed, I. Rashid, Md. G. R. Alam, A. K. M. Masum, And M. A. A. Dewan, "Shapley-Additive-Explanations-Based Factor Analysis For Dengue Severity Prediction Using Machine Learning," *J. Imaging*, Vol. 8, No. 9, P. 229, Aug. 2022, Doi: 10.3390/jimaging8090229.

BIOGRAPHY OF AUTHORS

	Surtiningsih, S.ST., M.Kes Biography: Surtiningsih is a lecturer at the Midwifery Study Program, Faculty of Health, Universitas Harapan Bangsa. With over 15 years of experience, my research focuses on maternal health, high-risk pregnancy, obstetric complications, and promotive and preventive midwifery care, integrating evidence-based complementary care. I also develop research related to early detection of complications, with an integration of clinical epidemiology and the development of a predictive model for obstetric complications. Email: surtiningsih@uhb.ac.id Orcid: https://orcid.org/0009-0005-4470-0097
	Linda Yanti, S.ST., M.Keb Biography: Linda Yanti is a lecturer at the Midwifery Study Program, Faculty of Health, Universitas Harapan Bangsa, with research interests in maternal health, high-risk pregnancy, obstetric complications, and community-based promotive and preventive midwifery care. Email: lindayanti@uhb.ac.id Orcid: https://orcid.org/0009-0005-8501-1562
	Hadi Jayusman, S. Kom., M.Ti Hadi Jayusman is a lecturer at the Information Systems Study Program, Universitas Harapan Bangsa. His research focuses on integrating artificial intelligence (AI) in community health, education, and digital solutions, with projects involving AI for disease diagnosis, cyberbullying mitigation, and the use of mobile systems and machine learning for medical prediction and community interventions. Email: hadijayusman@uhb.ac.id Orcid: https://orcid.org/0009-0003-5150-1630