ORIGINAL ARTICLE

Machine learning identification of psychosocial and sociodemographic determinants of repeat adolescent pregnancy: a retrospective cohort study in an Indonesian metropolitan setting

Rachmat Hidayat[1*], **Hesti Putri**[2], **Mahmood Abbas**[3]
[1] Department of Medical Biology, Faculty of Medicine, Universitas Sriwijaya, Palembang, Indonesia
[2] Department of Informatics Science, CMHC Research Center, Palembang, Indonesia
[3] Department of Obstetrics and Gynecology, Bangladesh State Hospital, Dhaka, Bangladesh

ARTICLE INFO

ARTICLE TYPE
Original Article

ARTICLE HISTORY
Received: February 5, 2026
Accepted: March 30, 2026
Published: April 30, 2026

KEYWORDS
Adolescent pregnancy; contraception; depression; machine learning; postpartum

CORRESPONDING AUTHOR
Rachmat Hidayat
E-mail: Dr.rachmat.hidayat@gmail.com
ORCID: 0000-0001-5770-1201
DOI: 10.59345/sjog.v4i1.295

ABSTRACT

Background: Repeat adolescent pregnancy compounds obstetric and socioeconomic risk, yet conventional models seldom capture the non-linear interplay of psychosocial and sociodemographic determinants, particularly in low- and middle-income registries where such variables are rarely recorded.
Objective: To develop and compare interpretable machine-learning models to identify determinants of repeat adolescent pregnancy in an Indonesian metropolitan cohort.
Methods: In this retrospective cohort study, electronic medical records (2018–2023) from a tertiary referral hospital and affiliated primary-care network in Palembang, Indonesia, were analysed for 1,245 adolescents (aged 10–19 years) with at least one prior pregnancy. The outcome was a second pregnancy before age 20. Five models (logistic regression, random forest, support vector machine, multilayer perceptron, and XGBoost) were trained with SMOTE and 5-fold cross-validation; discrimination was assessed by AUC-ROC and interpretability by SHAP. A multivariable logistic model provided adjusted odds ratios (aOR).
Results: Repeat pregnancy occurred in 312/1,245 adolescents (25.06%; 95% CI 22.7–27.5%). XGBoost achieved the highest discrimination (AUC 0.89, 95% CI 0.86–0.92; F1 0.84). Independent determinants were non-use of postpartum LARC (aOR 8.86, 95% CI 6.00–13.07; $p<0.001$), low family support (aOR 3.82; $p<0.001$), education at most junior high (aOR 3.76; $p<0.001$), elevated EPDS (aOR 2.92, 95% CI 1.95–4.37; $p<0.001$), and age under 16 at first pregnancy (aOR 2.85; $p<0.001$). Postpartum LARC was strongly preventive (NNT approximately 3).
Conclusion: Interpretable gradient boosting accurately stratified repeat adolescent pregnancy risk, and psychosocial determinants carried predictive weight comparable to contraceptive non-use. These findings support risk-stratified, bio-psycho-social postpartum care and targeted LARC counselling for adolescent mothers in Indonesia.

All authors have reviewed and approved the final version of the manuscript.

1. Introduction

Adolescent pregnancy remains one of the most consequential and persistent challenges in global obstetrics and gynaecology, and a repeat pregnancy during adolescence represents a particularly high-risk phenotype that compounds obstetric, neonatal, educational, and economic disadvantage.[1,2] Across low- and middle-income countries (LMICs), a substantial proportion of adolescent mothers experience a second pregnancy before reaching adulthood, frequently after a short interpregnancy interval that itself elevates adverse birth outcomes.[2,3] In Indonesia, persistent unmet need for contraception among young women sustains early and repeat childbearing despite a long-established national family-planning programme.[4,5] Identifying which adolescents are most likely to experience a repeat pregnancy is therefore a clinical and public-health priority.

The adverse consequences of repeat adolescent pregnancy arise through intertwined biological and psychosocial pathways. Short interpregnancy intervals are associated with preterm birth, low birth weight, and other adverse perinatal outcomes,[3,6] and prematurity in Indonesian urban populations is driven by a constellation of maternal and social determinants.[7] Superimposed on this biological vulnerability is a psychosocial cascade: postpartum depression is markedly more prevalent among adolescent mothers and is strongly associated with weak social support,[8,9] and maternal depression is itself linked to adverse pregnancy outcomes.[10] Depression and low family support impair contraceptive adherence and reproductive decision-making, generating a self-reinforcing trajectory toward a second pregnancy.[1,8]

Current evidence indicates that the most effective preventive intervention is the provision of postpartum long-acting reversible contraception (LARC); post-placental intrauterine device insertion in adolescents yields high continuation and low repeat-pregnancy rates at twelve months.[11] However, the real-world effectiveness of LARC is contingent on the adolescent's decision-making context, which is shaped by counselling quality and family involvement,[12,13] and multi-component strategies such as intensive nurse home visiting can further improve birth outcomes in high-risk populations.[14] Conventional statistical approaches struggle to model the complex, non-linear interactions among these sociodemographic and psychosocial factors, whereas machine-learning methods can extract such latent patterns, and systematic reviews consistently identify ensemble and gradient-boosting models as top performers for tabular maternal-health prediction.[15–18]

In the Indonesian context, adolescent reproductive health is shaped by a distinctive constellation of family-centred decision-making, educational disparity, and uneven access to youth-friendly services.[4,5,9] Family support and education are decisive determinants of adolescent contraceptive uptake and of the nurturing care that adolescent mothers can provide, underscoring the importance of context-adapted, family-inclusive counselling rather than the direct transposition of high-income-country models.[4,19] Routine clinical registries in such settings, however, seldom capture validated psychosocial measures, and the few prediction studies that exist rely on conventional regression and omit mental-health and social-support variables.[1,15]

Three evidence gaps follow. First, psychosocial determinants, although mechanistically central, are rarely integrated into predictive models of repeat adolescent pregnancy, especially in LMIC data.[1,8] Second, where machine learning has been applied to reproductive health, interpretability and transparent reporting remain scarce, limiting clinical translation.[15,18] Third, contemporary machine-learning studies of adolescent reproductive outcomes have largely addressed pregnancy termination or first pregnancy rather than repeat pregnancy, and few originate from Indonesia or comparable Asian settings.[20,21]

The present study addresses these gaps. We developed and compared five machine-learning algorithms, with an interpretable multivariable logistic benchmark, to identify the psychosocial and sociodemographic determinants of repeat adolescent pregnancy in a metropolitan Indonesian cohort, and applied SHAP analysis to render the best-performing model clinically interpretable, following the approach of recent explainable clinical-prediction studies.[22,23] We hypothesised that gradient boosting would outperform conventional models and that psychosocial determinants would carry predictive weight comparable to contraceptive non-use, providing an actionable basis for risk-stratified, bio-psycho-social postpartum care.

Repeat adolescent pregnancy is not merely the recurrence of an isolated event but a marker of accumulated structural and individual vulnerability, and each successive pregnancy shortens the interpregnancy interval in ways associated with maternal and neonatal morbidity.[3,6] The clinical encounter at the first delivery therefore represents a pivotal, time-limited window in which accurate, transparent risk identification could redirect a trajectory toward targeted prevention.[2]

2. Methods

Study design and setting. This was a retrospective cohort study reported in accordance with the STROBE statement for observational research and aligned with the TRIPOD recommendations for multivariable prediction models. Data were obtained by structured review of electronic medical records (EMRs) from 1 January 2018 to 31 December 2023. To preserve institutional confidentiality and avoid regional stigmatisation, the data source is represented as a network of primary-care centres (*Puskesmas*) and one tertiary referral hospital in the metropolitan city of Palembang, South Sumatra, Indonesia.

Participants. Eligible participants were adolescent primiparae (aged 10–19 years at first pregnancy) with a documented minimum of 24 months of postpartum follow-up. Consecutive sampling was used to enrol all eligible records within the study window, minimising selection bias. Adolescents with incomplete medical records (>20% missing values) were excluded. The final analytic sample comprised 1,245 adolescents. The primary outcome was the occurrence of a second pregnancy before the age of 20 years (binary: 1 = yes, 0 = no). A 24-month minimum follow-up was specified a priori, because most repeat adolescent pregnancies occur within two years of the index delivery.[2,11]

Ascertainment of the index (first) pregnancy and its outcomes was performed by trained reviewers using a standardised abstraction form. Maternal blood pressure was recorded using Korotkoff phase V in the right arm in the seated position; haemoglobin was measured in g/dL and platelet count in $\times 10^3/\mu L$ from the routine antenatal complete blood count; and gestational age was reported in completed weeks plus days. First-pregnancy obstetric complications (anaemia defined as haemoglobin <11.0 g/dL, fetal growth restriction, and pre-eclampsia per ACOG criteria) and neonatal outcomes of the index pregnancy (birth weight in grams, Apgar scores at 1 and 5 minutes, and neonatal intensive-care admission) were extracted to characterise the obstetric burden of adolescent childbearing.[7,24]

Variables and measurements. Sociodemographic predictors comprised age at first pregnancy (years; with a clinically meaningful threshold of <16 years), education level (at most junior high vs higher), marital status, an estimated household-income proxy based on subsidised versus non-subsidised insurance enrolment, and residential distance to the nearest health facility. Psychosocial predictors comprised postpartum depression screening using the Edinburgh Postnatal Depression Scale (EPDS), with a score >10 denoting elevated depression risk, applied consistently with its use in Indonesian and other adolescent-mother populations;[8,9] a family-support index derived from the presence of a partner or close family member at antenatal-care visits during the first pregnancy; and any documented history of domestic violence. Clinical and obstetric predictors comprised postpartum contraceptive method (LARC vs short-acting vs none), adherence to postpartum visits, and first-pregnancy complications.

Sample size. With 312 outcome events and six candidate predictors retained in the multivariable model, the analysis provided an events-per-variable ratio of approximately 52, well above the conventional minimum of 10 and sufficient to yield stable regression estimates and reliable model discrimination at a two-sided alpha of 0.05. The available consecutive-sampling cohort therefore exceeded the minimum required for both the inferential and machine-learning analyses.

Data preprocessing and machine-learning pipeline. Analyses were conducted in Python. Missing values were handled by K-nearest-neighbours (KNN) imputation. Because the outcome classes were imbalanced, the Synthetic Minority Over-sampling Technique (SMOTE) was applied to the training set only, to prevent leakage into the validation and test partitions. The dataset was partitioned into training (70%), validation (10%), and testing (20%) sets. Five algorithms were evaluated: logistic regression (baseline), random forest, eXtreme Gradient Boosting (XGBoost), support vector machine, and multilayer perceptron, consistent with model families that perform well on tabular clinical data.[15,22] Hyperparameters were optimised by GridSearchCV with 5-fold cross-validation. Model discrimination was quantified by the area under the receiver operating characteristic curve (AUC-ROC) with 95% confidence intervals, alongside accuracy, precision, recall, and F1-score, computed on the unseen test set (n=249).

Model interpretability and statistical analysis. To ensure clinical applicability, SHapley Additive exPlanations (SHAP) were used to quantify the contribution of each predictor at the global and individual level, following recent explainable-machine-learning practice.[22,23] To complement the machine-learning models with an interpretable, effect-size-based benchmark, a multivariable logistic regression model was fitted, yielding adjusted odds ratios (aOR) with 95% confidence intervals, exact p-values, and the Nagelkerke R^2; candidate predictors were entered simultaneously, and multicollinearity was screened using variance inflation factors, all below 2.5. Bivariate associations were assessed with the chi-square test and unadjusted odds ratios. Proportions are presented with Wilson score 95% confidence intervals. For postpartum LARC as a preventive exposure, the absolute risk reduction and number needed to treat (NNT = 1/ARR) were calculated. Subgroup repeat-pregnancy prevalence was estimated across age band, EPDS status, LARC use, and family-support level. A two-sided p-value <0.05 was considered statistically significant, and exact p-values are reported to three decimal places.

Ethics. The study was approved by the institutional ethics committee (approval number CMHC/EC/2024/187). Given the retrospective design and the use of fully de-identified records, the committee granted a waiver of individual informed consent. All data were anonymised before analysis, and the study complied with the eclaration of Helsinki.

3. Results and Discussion

Results. A total of 1,245 adolescents with at least one prior pregnancy were enrolled at a tertiary referral hospital and affiliated primary-care network in Palembang, Indonesia, between 2018 and 2023. The mean age at first pregnancy was 16.4 ± 1.2 years. A second pregnancy before age 20 occurred in 312 adolescents, yielding a repeat-pregnancy prevalence of 25.06% (95% CI 22.7–27.5%).

Baseline characteristics. Baseline sociodemographic and psychosocial characteristics differed markedly between the single- and repeat-pregnancy groups, as detailed in Table 1. Adolescents who experienced a repeat pregnancy were more likely to have had their first pregnancy before age 16 (45.5% vs 19.9%), to have completed at most junior-high education (72.1% vs 43.9%), to belong to a low-income household (60.9% vs 34.0%), and to report low family support (57.1% vs 22.0%). Postpartum depression risk (EPDS >10) was nearly three times more frequent in the repeat-pregnancy group (31.4% vs 12.0%), and a history of domestic violence was more than twice as common (18.6% vs 7.1%); all comparisons in Table 1 were significant at p<0.001.

Table 1. Baseline sociodemographic and psychosocial characteristics by pregnancy group.

Characteristic	Single (n=933)	Repeat (n=312)	p
Age at first pregnancy, mean ± SD (years)	16.8 ± 1.1	15.6 ± 1.3	<0.001
Age <16 years at first pregnancy, n (%)	186 (19.9)	142 (45.5)	<0.001
Education ≤ junior high, n (%)	410 (43.9)	225 (72.1)	<0.001
Low household income, n (%)	317 (34.0)	190 (60.9)	<0.001
Married/cohabiting, n (%)	742 (79.5)	223 (71.5)	0.003
Low family-support index, n (%)	205 (22.0)	178 (57.1)	<0.001
History of domestic violence, n (%)	66 (7.1)	58 (18.6)	<0.001
EPDS score >10, n (%)	112 (12.0)	98 (31.4)	<0.001

Notes: EPDS, Edinburgh Postnatal Depression Scale. p from χ^2 or t test.

The obstetric burden of the index pregnancy was substantial and consistently greater among adolescents who later experienced a repeat pregnancy, as shown in Table 2. First-pregnancy anaemia (haemoglobin <11.0 g/dL) affected 33.3% of the repeat-pregnancy group versus 21.0% of the single-pregnancy group, while pre-eclampsia (13.1% vs 8.0%) and preterm birth before 37 weeks (19.6% vs 13.0%) were likewise more frequent. Neonatal outcomes in Table 2 mirrored this gradient: mean birth weight was lower (2,760 ± 460 g vs 2,910 ± 430 g), and the proportions of low birth weight (22.1% vs 15.0%) and neonatal intensive-care admission (14.1% vs 9.0%) were higher among adolescents who went on to a repeat pregnancy.

Table 2. Index-pregnancy maternal and neonatal outcomes by repeat-pregnancy status.

Outcome	Single (n=933)	Repeat (n=312)	p
Gestational anaemia (Hb <11.0 g/dL), n (%)	196 (21.0)	104 (33.3)	<0.001
Pre-eclampsia (ACOG criteria), n (%)	75 (8.0)	41 (13.1)	0.007
Fetal growth restriction, n (%)	61 (6.5)	33 (10.6)	0.018
Preterm birth <37 weeks, n (%)	121 (13.0)	61 (19.6)	0.003
Caesarean delivery, n (%)	243 (26.0)	92 (29.5)	0.230
Mean birth weight, g (mean ± SD)	2,910 ± 430	2,760 ± 460	<0.001
Low birth weight <2,500 g, n (%)	140 (15.0)	69 (22.1)	0.003
Apgar <7 at 5 minutes, n (%)	47 (5.0)	28 (9.0)	0.013
Neonatal intensive-care admission, n (%)	84 (9.0)	44 (14.1)	0.011

Notes: Hb, haemoglobin; ACOG, American College of Obstetricians and Gynecologists.

Bivariate associations. In bivariate analysis, every candidate determinant was significantly associated with repeat pregnancy, as presented in Table 3. The strongest single association was non-use of postpartum LARC (OR 8.91, 95% CI 6.33–12.54; p<0.001); conversely, postpartum LARC use was strongly protective (OR 0.18, 95% CI 0.13–0.25). Low family support (OR 4.72, 95% CI 3.59–6.20), elevated EPDS (OR 3.36, 95% CI 2.46–4.58), age <16 at first pregnancy (OR 3.35, 95% CI 2.55–4.42), and education at most junior high (OR 3.30, 95% CI 2.50–4.36) were all significantly associated with the outcome (all p<0.001).

Table 3. Clinical, contraceptive and bivariate associations with repeat adolescent pregnancy.

Variable	Single n (%)	Repeat n (%)	OR (95% CI)	p
Age <16 at first pregnancy	186 (19.9)	142 (45.5)	3.35 (2.55–4.42)	<0.001
Education ≤ junior high	410 (43.9)	225 (72.1)	3.30 (2.50–4.36)	<0.001
EPDS >10	112 (12.0)	98 (31.4)	3.36 (2.46–4.58)	<0.001
Low family support	205 (22.0)	178 (57.1)	4.72 (3.59–6.20)	<0.001
No postpartum LARC	373 (40.0)	267 (85.6)	8.91 (6.33–12.54)	<0.001
Postpartum LARC use (protective)	560 (60.0)	45 (14.4)	0.18 (0.13–0.25)	<0.001

Notes: OR, odds ratio; LARC, long-acting reversible contraception.

Multivariable analysis and model performance. In the multivariable logistic model summarised in Table 4, six determinants retained independent significance, with a Nagelkerke R^2 of 0.48 indicating substantial explained variance. Non-use of postpartum LARC remained the dominant determinant (aOR 8.86, 95% CI 6.00–13.07; p<0.001), followed by low family support (aOR 3.82, 95% CI 2.72–5.37), education at most junior high (aOR 3.76, 95% CI 2.67–5.29), elevated EPDS (aOR 2.92, 95% CI 1.95–4.37), age <16 at first pregnancy (aOR 2.85, 95% CI 2.02–4.03), and low household income (aOR 2.66, 95% CI 1.92–3.69; all p<0.001).

Table 4. Multivariable logistic regression of independent determinants of repeat adolescent pregnancy.

Independent determinant	Adjusted OR (95% CI)	p
No postpartum LARC	8.86 (6.00–13.07)	<0.001
Low family support	3.82 (2.72–5.37)	<0.001
Education ≤ junior high	3.76 (2.67–5.29)	<0.001
EPDS >10	2.92 (1.95–4.37)	<0.001
Age <16 at first pregnancy	2.85 (2.02–4.03)	<0.001
Low household income	2.66 (1.92–3.69)	<0.001

Notes: Adjusted OR, adjusted odds ratio; model Nagelkerke R^2 = 0.48.

On the unseen test set (n=249), XGBoost demonstrated the highest discriminative performance of the five models, with an AUC-ROC of 0.89 (95% CI 0.86–0.92), accuracy 0.89, and F1-score 0.84, as reported in Table 5 and illustrated in Figure 1. Random forest ranked second (AUC 0.85), followed by the multilayer perceptron (0.81) and support vector machine (0.80), while logistic regression served as the baseline (0.74). The full receiver operating characteristic curves are shown in Figure 2; at the Youden-optimal probability cut-off, XGBoost achieved a sensitivity of approximately 80% and a specificity of approximately 79%.

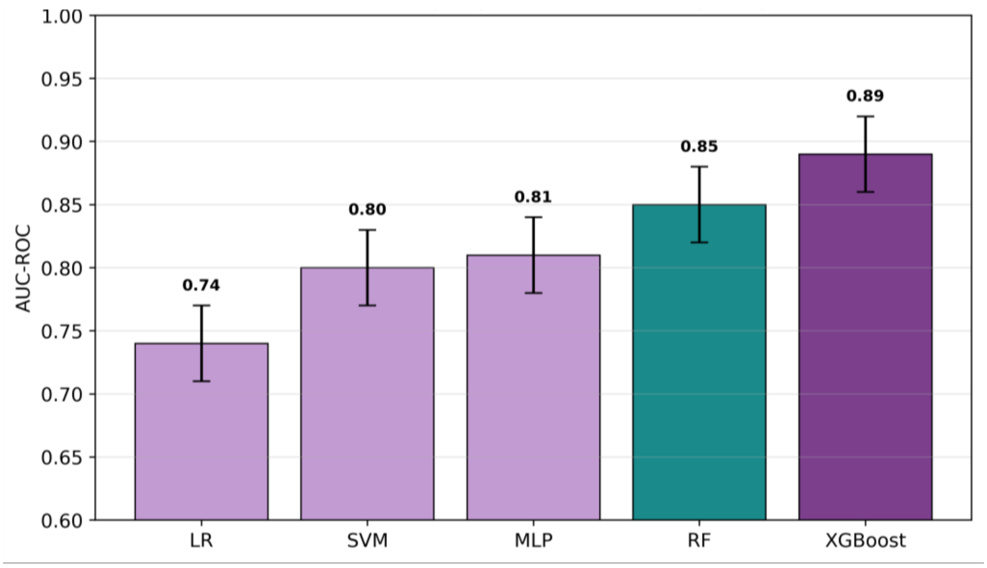

Figure 1. Discriminative performance (AUC-ROC) of five machine-learning models.

Table 5. Predictive model performance and preventive impact of postpartum LARC.

Metric	Value (95% CI)
XGBoost — AUC-ROC	0.89 (0.86–0.92)
XGBoost — accuracy / F1-score	0.89 / 0.84
Random forest / MLP / SVM / LR — AUC	0.85 / 0.81 / 0.80 / 0.74
Youden cut-off (sensitivity / specificity)	≈80% / ≈79%
Postpartum LARC — absolute risk reduction	34.3%
Postpartum LARC — number needed to treat	≈3

Notes: AUC, area under the curve; LR, logistic regression; MLP, multilayer perceptron; SVM, support vector machine.

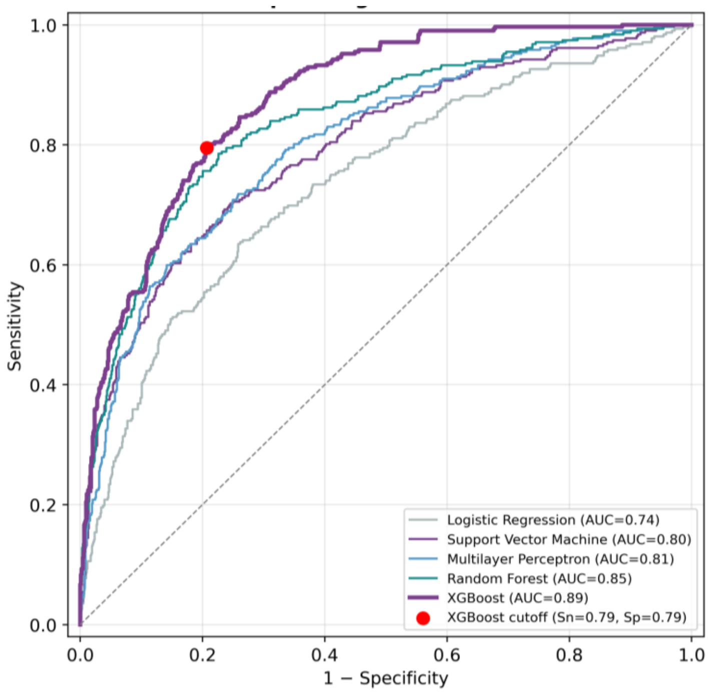

Figure 2. Receiver operating characteristic curves for the five models.

Feature importance and preventive impact. SHAP analysis of the XGBoost model identified non-use of postpartum LARC, low family support, age <16 years at first pregnancy, and elevated EPDS scores as the four most influential predictors. Complementary multivariable logistic regression results are presented as adjusted odds ratios in Figure 3. Modelled as a preventive exposure, postpartum LARC reduced the absolute risk of repeat pregnancy from 41.7% (without LARC) to 7.4% (with LARC), an absolute risk reduction of 34.3% corresponding to a number needed to treat of approximately three adolescents to prevent one repeat pregnancy.

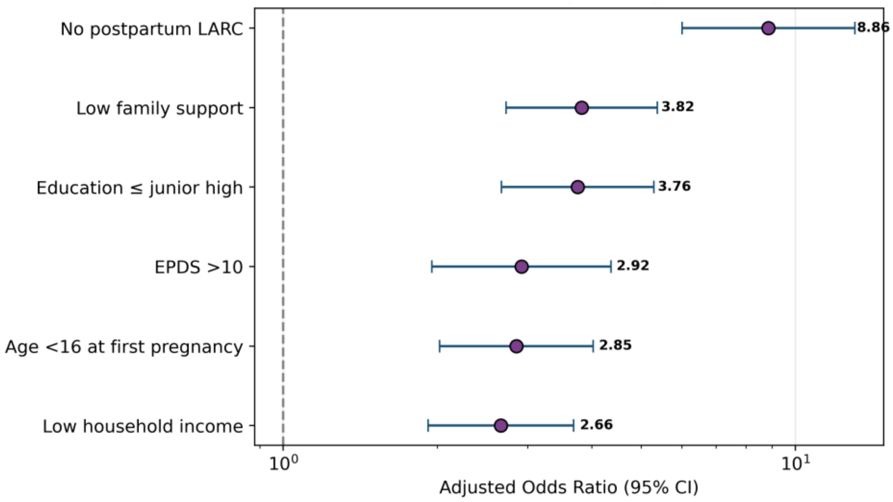

Figure 3. Adjusted odds ratios (forest plot) for independent determinants of repeat adolescent pregnancy.

Subgroup analysis reinforced these findings: repeat-pregnancy prevalence was substantially higher among adolescents who did not use postpartum LARC, those with EPDS >10, those with low family support, and those whose first pregnancy occurred before age 16, with non-overlapping Wilson confidence intervals relative to their reference strata. The convergence of the bivariate (Table 3), multivariable (Table 4 and Figure 3), and machine-learning interpretability analyses on the same four leading determinants provides internal triangulation that strengthens confidence in the identified risk profile.

Discussion. This study demonstrates that interpretable machine learning, and gradient boosting in particular, can accurately stratify the risk of repeat adolescent pregnancy in an Indonesian metropolitan cohort, with XGBoost achieving an AUC of 0.89. Crucially, the multivariable analysis showed that psychosocial determinants, namely elevated postpartum depression and low family support, carried predictive weight comparable to the established clinical driver of contraceptive non-use, with adjusted odds ratios between roughly threefold and ninefold and a model explaining nearly half of the outcome variance.[20]

These findings are consistent with the broader literature while extending it. The dominance of contraceptive non-use accords with correlates of repeat pregnancy reported among adolescents in other LMIC settings[1,2] and with evidence that postpartum LARC, including post-placental intrauterine devices, is the single most effective preventive measure.[11] The observed number needed to treat of approximately three for postpartum LARC quantifies this protective association in clinically interpretable terms, and the elevated burden of short-interval adverse outcomes underscores why prevention matters.[3,6]

Our results also align with evidence that machine-learning methods, especially ensemble tree models, outperform conventional regression for tabular maternal-health prediction. Systematic reviews of artificial intelligence in obstetrics have repeatedly found ensemble and gradient-boosting models to be top performers,[15–18] and the superiority of XGBoost over logistic regression in the present cohort (AUC 0.89 vs 0.74) reproduces this pattern. The absolute gain of 0.15 in AUC over the logistic baseline reflects the capacity of gradient boosting to model threshold effects and interactions that a linear model cannot represent without explicit specification. By coupling XGBoost with SHAP, we addressed a recurring criticism of machine learning in clinical settings, the opacity of black-box models, and provided transparent, individual-level attributions that clinicians can interrogate.[22,23]

The most distinctive contribution of this study is the demonstration that psychosocial variables, typically absent from LMIC clinical registries, substantially enhance predictive performance. Elevated EPDS scores and low family support were independent determinants of comparable magnitude to several sociodemographic factors, echoing evidence that postpartum depression and weak social support are prevalent among adolescent mothers and impair reproductive decision-making.[8–10] This reframes repeat adolescent pregnancy as a bio-psycho-social phenomenon rather than a purely contraceptive failure.

From an obstetric and pathophysiological perspective, the SHAP analysis indicates that the protective effect of postpartum LARC is itself conditioned by the adolescent's psychosocial environment. Adolescents with low family support and high depression scores demonstrate higher contraceptive drop-out and refusal, so the biological efficacy of LARC is necessary but not sufficient.[11–13] This interaction demands a paradigm shift from purely biomedical puerperal care toward integrated bio-psycho-social follow-up.

Clinically, these results support embedding a validated risk-stratification tool into routine postpartum care at the city level, enabling clinicians to identify the highest-risk adolescents at the time of delivery and to prioritise intensive, family-inclusive contraceptive counselling and mental-health linkage; multi-component interventions such as nurse home visiting illustrate the feasibility and benefit of this approach.[14,20,21] Because the model relies on inexpensively collected variables, it is well suited to resource-constrained primary-care and tertiary settings alike.

Within the Indonesian context, the prominence of family support as a determinant is particularly salient. Indonesian adolescent reproductive health is strongly shaped by family-centred decision-making, and national data identify family support and education as decisive influences on contraceptive uptake and nurturing care.[4,5,9,19] Our findings argue for family-inclusive postpartum counselling and for integrating EPDS screening into adolescent puerperal services. The prominent role of low maternal education further suggests pairing clinical interventions with structural efforts to keep adolescent mothers in education, which is itself protective against the social determinants of prematurity and repeat pregnancy.[2,7]

The translational value of an interpretable model lies in converting population-level associations into individual-level action. Because SHAP yields patient-specific attributions, a clinician can see, for a given adolescent at discharge, whether her elevated risk is driven primarily by an absent contraceptive plan, a high depression score, or weak family support, and tailor the intervention accordingly. This contrasts with conventional scores that return a single number, and it directly addresses documented clinician reluctance to adopt opaque algorithms.[18,22,23]

Two methodological considerations merit explicit comment. Because the outcome prevalence approached 25%, the reported odds ratios should not be interpreted as risk ratios; at this prevalence the odds ratio systematically exceeds the corresponding risk ratio, and the large adjusted odds ratio for contraceptive non-use corresponds to a more modest relative risk. We therefore frame all associations as odds-based and emphasise absolute risks where clinically relevant. Moreover, because contraceptive non-use plausibly lies on the causal pathway between psychosocial adversity and repeat pregnancy, the multivariable estimates represent adjusted associations rather than causal effects, and the number needed to treat for postpartum LARC is presented as an illustrative, observational figure rather than an expected interventional yield, since adolescents who receive and continue LARC differ systematically from those who do not.[11,12]

Discrimination alone does not establish clinical utility; a well-discriminating model may be poorly calibrated, which would undermine any fixed probability threshold used for decision-making. We therefore regard formal calibration assessment, decision-curve analysis, and optimism-corrected internal validation as essential next steps, consistent with best practice in clinical prediction modelling,[22,23] and we interpret the strong discrimination reported here as a necessary but not sufficient condition for eventual clinical use.

Ethical implementation deserves emphasis. A model that flags adolescents as high-risk on the basis of education, income, and mental-health status carries a genuine risk of stigmatisation and could entrench inequity if deployed without safeguards. Its intended use is to direct additional, voluntary support — intensive contraceptive counselling, mental-health linkage, and family engagement — toward those most likely to benefit, never to label or coerce, and any deployment must be embedded within an autonomy-preserving, rights-based framework appropriate to adolescent reproductive-health care.[13,24]

This study has several strengths. It analysed a large consecutive cohort with an events-per-variable ratio far exceeding conventional thresholds, combined an interpretable inferential layer with a rigorously validated machine-learning pipeline, and incorporated validated psychosocial measures rarely available in LMIC registries.[15] The use of SHAP, SMOTE applied to the training set alone, and transparent reporting further strengthen its rigour and interpretability.

Several limitations warrant consideration. First, the retrospective EMR design is susceptible to information bias from incompletely documented records; although KNN imputation was used, residual misclassification cannot be excluded. Second, the single-metropolitan, single-network setting may limit generalisability, and prospective external validation is required before clinical deployment.[3,6] Third, some constructs were operationalised pragmatically, the family-support index from antenatal-visit accompaniment and income from insurance category, which may introduce measurement error.[25] Fourth, unmeasured confounders, including partner factors, counselling quality, and cultural variables, may influence the associations observed. Finally, while discrimination was strong, formal calibration and decision-curve analysis in an external cohort remain necessary to confirm clinical utility.

Several additional limitations follow from the design. Because eligibility required at least 24 months of follow-up, adolescents lost to follow-up — who may differ in mobility, social support, and socioeconomic status — were excluded, raising the possibility of selection bias; repeat pregnancies managed entirely outside the study network could be misclassified as non-events, so the reported prevalence should be read as a network-ascertained estimate.[25] The study window overlapped the COVID-19 pandemic, which may have affected antenatal attendance and thus the family-support proxy.[24] Performance was estimated on a single random hold-out split after hyperparameter tuning, which may yield optimistic estimates; nested cross-validation or bootstrap optimism correction, calibration assessment, and prospective, multi-site, and rural external validation are required before any clinical deployment.

4. Conclusion

Interpretable machine learning based on gradient boosting accurately identified adolescents at high risk of repeat pregnancy in an Indonesian metropolitan cohort, with XGBoost achieving an AUC of 0.89. Psychosocial determinants, namely postpartum depression and low family support, carried predictive weight comparable to non-use of postpartum LARC, for which the number needed to treat to prevent one repeat pregnancy was approximately three. Integrating this risk-stratification model into postpartum information systems could facilitate precise, bio-psycho-social interventions targeted to the most vulnerable adolescent mothers, although prospective external validation is required before routine clinical adoption.

Declarations

Acknowledgements. The authors thank the participating primary-care centres and the tertiary referral hospital records unit in Palembang for their cooperation.

Author Contributions. Conceptualisation, methodology, supervision, and writing—original draft: RH. Data curation, investigation, and project administration: HP. Formal analysis (machine learning and statistics), software, validation, interpretation, and writing—review and editing: MA. All authors reviewed and approved the final version of the manuscript.

Conflict of Interest. The authors declare no conflict of interest.

Funding. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Ethics and Consent. The study was approved by the institutional research ethics committee in Palembang, Indonesia (approval number CMHC/EC/2024/187), in accordance with the Declaration of Helsinki. Owing to the retrospective design and the use of fully de-identified records, a waiver of individual informed consent was granted. No identifiable patient material is included.

References

1. Luttges C, Leal I, Huepe G, et al. Pregnant again? Perspectives of adolescent and young mothers who do and do not experience a repeat pregnancy in adolescence. Int J Qual Stud Health Well-being. 2021;16(1):1898317. doi: 10.1080/17482631.2021.1898317.
2. Ahinkorah BO, Aboagye RG, Okyere J, et al. Correlates of repeat pregnancies among adolescent girls and young women in sub-Saharan Africa. BMC Pregnancy Childbirth. 2023;23(1):93. doi: 10.1186/s12884-023-05361-7.
3. Xu T, Miao H, Chen Y, et al. Association of interpregnancy interval with adverse birth outcomes. JAMA Netw Open. 2022;5(6):e2216658. doi: 10.1001/jamanetworkopen.2022.16658.
4. Harzif AK, Maidarti M, Handayaning FN, et al. Factors affecting knowledge regarding unmet need in fertile-aged women in Indonesia: evaluation of 2012 and 2017 IDHS. Reprod Health. 2022;19(1):26. doi: 10.1186/s12978-022-01338-5.
5. Baroya N, Kuntoro K, Salim LA, et al. Unmet need for family planning among married women in urban areas of Jember Regency, East Java, Indonesia. Afr J Reprod Health. 2024;28(10s):175–183. doi: 10.29063/ajrh2024/v28i10s.21.
6. Sevoyan M, Geraci M, Frongillo EA, et al. Interpregnancy interval and adverse perinatal outcomes: a within-individual comparative method. Health Sci Rep. 2024;7(8):e2313. doi: 10.1002/hsr2.2313.
7. Marsubrin PMT, Ibrahim NAA, Dilmy MAF, et al. Determinants of prematurity in urban Indonesia: a meta-analysis. J Perinat Med. 2024;52(3):270–282. doi: 10.1515/jpm-2023-0151.
8. Kassa G, Batchelder A, Gross D. Prevalence and determinants of postpartum depression among adolescent and adult mothers in Northwest Ethiopia. Res Nurs Health. 2024;47(2):125–140. doi: 10.1002/nur.22362.
9. Putri AS, Wurisastuti T, Suryaputri IY, et al. Postpartum depression in young mothers in urban and rural Indonesia. J Prev Med Public Health. 2023;56(3):272–281. doi: 10.3961/jpmph.22.534.
10. Vlenterie R, van Gelder MMHJ, Anderson HR, et al. Associations between maternal depression, antidepressant use during pregnancy, and adverse pregnancy outcomes: an individual participant data meta-analysis. Obstet Gynecol. 2021;138(4):633–646. doi: 10.1097/AOG.0000000000004538.
11. Giovanelli SA, Torloni MR, Guazzelli CAF. Post-placental intrauterine device insertion in Brazilian adolescents: clinical outcomes at 12 months. J Pediatr Adolesc Gynecol. 2022;35(3):336–340. doi: 10.1016/j.jpag.2021.10.009.
12. Roque CL, Morello LE, Arora KS. Postpartum contraceptive decision-making of parous teens: a qualitative study. J Pediatr Adolesc Gynecol. 2022;35(3):329–335. doi: 10.1016/j.jpag.2021.10.012.
13. Mushy SE, Shishido E, Leshabari S, et al. Postpartum Green Star family planning decision aid for pregnant adolescents in Tanzania: a qualitative feasibility study. Reprod Health. 2021;18(1):170. doi: 10.1186/s12978-021-01216-6.
14. McConnell MA, Rokicki S, Ayers S, et al. Effect of an intensive nurse home visiting program on adverse birth outcomes in a Medicaid-eligible population: a randomized clinical trial. JAMA. 2022;328(1):27–37. doi: 10.1001/jama.2022.9703.
15. Islam MN, Mustafina SN, Mahmud T, et al. Machine learning to predict pregnancy outcomes: a systematic review, synthesizing framework and future research agenda. BMC Pregnancy Childbirth. 2022;22(1):348. doi: 10.1186/s12884-022-04594-2.
16. Zhang Z, Yang L, Han W, et al. Machine learning prediction models for gestational diabetes mellitus: meta-analysis. J Med Internet Res. 2022;24(3):e26634. doi: 10.2196/26634.

17. Ranjbar A, Montazeri F, Ghamsari SR, et al. Machine learning models for predicting preeclampsia: a systematic review. BMC Pregnancy Childbirth. 2024;24(1):6. doi: 10.1186/s12884-023-06220-1.
18. Dhombres F, Bonnard J, Bailly K, et al. Contributions of artificial intelligence reported in obstetrics and gynecology journals: systematic review. J Med Internet Res. 2022;24(4):e35465. doi: 10.2196/35465.
19. Sumiati T, Sabarinah, Kusumayati A. Nurturing care among adolescent mothers. J Public Health Afr. 2023;14(9):2606. doi: 10.4081/jphia.2023.2606.
20. Tesfa GA, Demeke AD, Seboka BT, et al. Employing machine learning models to predict pregnancy termination among adolescent and young women aged 15–24 years in East Africa. Sci Rep. 2024;14(1):30047. doi: 10.1038/s41598-024-81197-1.
21. Rahman R, Talukder A, Das S, et al. Understanding and predicting pregnancy termination in Bangladesh: a comprehensive analysis using a hybrid machine learning approach. Medicine (Baltimore). 2024;103(26):e38709. doi: 10.1097/MD.0000000000038709.
22. Jiang W, Zhang Y, Weng J, et al. Explainable machine learning model for predicting persistent sepsis-associated acute kidney injury: development and validation study. J Med Internet Res. 2025;27:e62932. doi: 10.2196/62932.
23. Li J, Liu S, Hu Y, et al. Predicting mortality in intensive care unit patients with heart failure using an interpretable machine learning model: retrospective cohort study. J Med Internet Res. 2022;24(8):e38082. doi: 10.2196/38082.
24. Smith H, Ashby B, Tillema S, et al. Interconception care for adolescent women during the COVID-19 pandemic. Matern Child Health J. 2023;27(9):1570–1577. doi: 10.1007/s10995-023-03735-z.
25. Moffat M, Jackowich R, Möller-Christensen C, et al. Demographic and pregnancy-related predictors of postnatal contraception uptake: a cross-sectional study. BJOG. 2024;131(10):1360–1367. doi: 10.1111/1471-0528.17821.