An Interpretable, Equity-Audited Machine Learning Framework for Prenatal Prediction of Low Birth Weight Integrating Social Determinants in Honduras

Karen Oliva[1]; Angely Suazo[2]; Diana Zelaya[3]; Oneyda Zelaya[4]; Yenny Velasquez[5]; Antonieta Zuniga[6]
[1,2,5]Universidad Nacional Autónoma de Honduras, Honduras, *karen.oliva@unah.edu.hn, angely.suazo@unah.edu.hn, yenny.velasquez@unah.edu.hn*
[3,4]Policia Nacional de Honduras, Honduras, *lizethzelaya2010@gmail.com, oneydaespinoza1996@gmail.com*
[6]Ministerio Público de Honduras, Honduras, *antonietazuniga2311@yahoo.com*

Abstract– ***This study aimed to , internally validate, and operationalize an implementation-oriented framework for the prenatal prediction of low birth weight in Honduras by integrating routinely available maternal variables with social determinants of health. Models were trained on microdata from the Honduras ENDESA/MICS 2019 survey (N=5765 singleton live births with valid birthweight measurements), using maternal and household indicators suitable for primary care settings, including maternal age, educational attainment, wealth quintile, area of residence, anthropometric status, anemia, gestational age, and continuity of antenatal care. Logistic regression, Random Forest, and XGBoost classifiers were compared according to discrimination, calibration, sensitivity, and clinical utility metrics. Internal validation used stratified five-fold cross-validation, and 95% confidence intervals for the AUC ROC were obtained by bootstrapping (2,000 iterations). The prevalence of low birth weight was 9.0%, with higher estimates observed among mothers with low educational attainment, adolescent pregnancies, insufficient antenatal contacts, and greater social vulnerability. XGBoost showed the best overall performance, with an AUC ROC of 0.768 (95% CI: 0.73–0.81), sensitivity of 74.2%, and a Brier score of 0.075. SHAP-based interpretation identified gestational age, antenatal contacts, and maternal schooling as the most influential predictors. The contribution is not a novel algorithm but an implementation framework that couples the model with subgroup equity auditing, probability calibration, and an offline operating mode tailored to Honduran primary care. These findings suggest that a lightweight and interpretable predictive model could strengthen prenatal risk stratification in resource-constrained settings. Its integration into the Honduran electronic health record ecosystem may support early counseling, nutritional supplementation, prioritization of follow-up, and territorial monitoring. Operational feasibility with the Honduran Ministry of Health and primary care teams has not yet been assessed..***

Keywords-- ***low birth weight, social determinants of health, machine learning, Honduras, prenatal care.***

Un marco de aprendizaje automático interpretable y auditado por equidad para la predicción prenatal de bajo peso al nacer, integrando determinantes sociales en Honduras

Karen Oliva[1]; Angely Suazo[2]; Diana Zelaya[3]; Oneyda Zelaya[4]; Yenny Velasquez[5]; Antonieta Zuniga[6]
[1,2,5]Universidad Nacional Autónoma de Honduras, Honduras, *karen.oliva@unah.edu.hn, angely.suazo@unah.edu.hn, yenny.velasquez@unah.edu.hn*
[3,4]Policia Nacional de Honduras, Honduras, *lizethzelaya2010@gmail.com, oneydaespinoza1996@gmail.com*
[6]Ministerio Público de Honduras, Honduras, *antonietazuniga2311@yahoo.com*

Resumen– ***El objetivo de este estudio fue desarrollar validar internamente y operacionalizar un marco de implementación para la predicción prenatal del bajo peso al nacer en Honduras, integrando variables maternas rutinarias y determinantes sociales de la salud. Se realizó un estudio metodológico de desarrollo y validación interna de modelos predictivos a partir de datos maternos, clínicos y sociales procedentes de los microdatos de la encuesta ENDESA/MICS Honduras 2019, con 5,765 registros de nacimiento con peso válido. El bajo peso al nacer se definió como peso inferior a 2,500 gramos. Se compararon modelos de regresión logística penalizada, Random Forest y XGBoost, incorporando control del desbalance de clases, validación cruzada estratificada de cinco partes, calibración, estimación de intervalos de confianza del 95% para el AUC ROC mediante arranque con 2,000 iteraciones e interpretabilidad mediante valores SHAP. La prevalencia basal de bajo peso al nacer fue de 9.0%, con mayor frecuencia en madres con baja escolaridad, gestación adolescente, menor continuidad prenatal y vulnerabilidad socioeconómica. XGBoost mostró el mejor desempeño global, con AUC ROC de 0.768 (IC 95%: 0.73–0.81), sensibilidad de 74.2%, especificidad de 76.9% y Brier score de 0.075. Las variables con mayor contribución predictiva fueron edad gestacional, número de controles prenatales y escolaridad materna. El aporte no reside en un algoritmo nuevo, sino en un marco de implementación que acopla el modelo con auditoría de equidad por subgrupos, calibración de probabilidades y un modo de operación fuera de línea, adaptado al primer nivel de atención hondureño. Estos hallazgos sugieren que un modelo predictivo simple, interpretable y de bajo costo puede fortalecer la estratificación prenatal del riesgo, orientar intervenciones focalizadas y apoyar la planificación territorial en servicios de atención primaria en Honduras. La factibilidad operativa con la Secretaría de Salud y los equipos de atención primaria aún no se ha evaluado.***

Palabras clave-- bajo peso al nacer, determinantes sociales, aprendizaje automático, Honduras, atención prenatal.

I. Introducción

El bajo peso al nacer (BPN) constituye un indicador crítico de salud materna, neonatal y social. La Organización Mundial de la Salud (OMS) lo define como un peso al nacimiento inferior a 2,500 gramos, medido preferiblemente en las primeras horas de vida [1]. Esta condición resume una trayectoria gestacional afectada por prematuridad, restricción del crecimiento intrauterino o ambas, y se asocia con mortalidad neonatal, retraso del crecimiento, deterioro cognitivo y mayor riesgo de enfermedades crónicas en etapas posteriores de la vida [1], [2]. A escala poblacional, el BPN también refleja la calidad del cuidado prenatal, la nutrición materna, la seguridad alimentaria del hogar y la capacidad del sistema sanitario para identificar oportunamente embarazos vulnerables.

Las estimaciones internacionales del Fondo de las Naciones Unidas para la Infancia (UNICEF) y OMS muestran que más de 20 millones de recién nacidos presentan BPN cada año, lo que equivale aproximadamente a uno de cada siete nacimientos vivos [2]. Aunque las cifras nacionales suelen reportarse como promedios, su distribución no es homogénea. En países de ingresos bajos y medianos, las desigualdades territoriales y socioeconómicas generan núcleos de riesgo que pueden pasar inadvertidos cuando el análisis se limita a indicadores agregados. Honduras comparte este desafío, pues combina dispersión geográfica, brechas de acceso, embarazo adolescente, inseguridad alimentaria en hogares vulnerables y heterogeneidad en la continuidad del control prenatal.

En el contexto latinoamericano, la prevalencia de BPN presenta variaciones significativas asociadas al entorno social urbano. Estudios multicéntricos en 360 ciudades de la región han documentado que las desigualdades socioeconómicas a nivel de vecindario se asocian con diferencias sustanciales en la prevalencia de BPN, incluso después de ajustar por factores individuales [3]. Datos recientes de encuestas demográficas y de salud en 44 países de ingreso bajo y mediano estiman una prevalencia regional del 11.5% para América Latina y el Caribe, superior al promedio africano (9.5%), aunque inferior al asiático (16%) [4]. Brasil ha reportado asociaciones entre BPN, determinantes socioeconómicos y obesidad peripuberal, lo que confirma que el BPN no es solo un evento neonatal sino

un predictor de riesgo metabólico a lo largo de la vida [5]. Estos hallazgos refuerzan la necesidad de modelos predictivos que capturen la interacción entre vulnerabilidad social y riesgo clínico, particularmente en sistemas de salud con recursos limitados como el hondureño.

El enfoque clínico tradicional ha priorizado factores biomédicos, como prematuridad, anemia, hipertensión gestacional, talla materna baja, índice de masa corporal pregestacional deficiente e infecciones no tratadas. Sin embargo, la epidemiología social ha demostrado que estos factores no actúan de forma aislada. La escolaridad materna, el ingreso, el área de residencia, la distancia al establecimiento, la disponibilidad de transporte y la capacidad de sostener controles prenatales condicionan la exposición al riesgo y la oportunidad de recibir intervenciones preventivas [6], [7]. Por tanto, un sistema de predicción útil para Honduras debe integrar determinantes sociales de la salud sin perder simplicidad operativa.

Las revisiones sistemáticas recientes sobre predicción de peso al nacer mediante aprendizaje automático han confirmado que los modelos basados en ensambles de árboles (XGBoost, LightGBM, Random Forest) superan consistentemente a los modelos logísticos convencionales cuando las relaciones entre predictores y desenlace son no lineales o presentan interacciones de orden superior [8], [9]. Sin embargo, estos estudios también señalan que la mayoría de los modelos publicados omiten la calibración de probabilidades y la evaluación de equidad por subgrupos, dos aspectos que el presente trabajo aborda explícitamente. Conviene precisar, no obstante, que los ensambles de árboles ya se han aplicado a la predicción del BPN en otros contextos, en algunos casos acompañados de técnicas de interpretabilidad [9], [10], [11], [12]; el algoritmo, por sí mismo, no constituye la novedad de este trabajo. La selección de XGBoost se fundamenta en su capacidad demostrada para datos tabulares con desbalance de clases, su compatibilidad con métodos de explicabilidad post-hoc como SHAP (explicaciones aditivas basadas en valores de Shapley), y su eficiencia computacional para despliegue en infraestructuras de bajo costo [10].

La oportunidad tecnológica reside en transformar datos ya disponibles en una herramienta de estratificación prenatal. Los modelos de aprendizaje automático pueden capturar interacciones no lineales entre variables maternas, clínicas y sociales, mientras que los métodos de explicación local permiten presentar predictores comprensibles para los equipos de salud. La brecha que este trabajo busca cerrar es, por tanto, de traslación. En Honduras, la clasificación del riesgo obstétrico que se aplica durante el control prenatal se apoya en listas de factores clínico-obstétricos que no estiman una probabilidad individual calibrada ni incorporan de forma explícita los determinantes sociales. Este trabajo presenta un marco de implementación que acopla un modelo interpretable, entrenado con datos nacionales, con calibración de probabilidades, auditoría de equidad por subgrupos, un modo de operación fuera de línea y una guía de adopción para orientar decisiones en el primer nivel de atención, priorizando acciones factibles como consejería nutricional, suplementación, seguimiento comunitario y controles adicionales.

II. Materiales y métodos

A. Diseño del estudio

Se desarrolló un estudio con orientación aplicada, centrada en el desarrollo, evaluación interna e implementación de un modelo predictivo para estimar el riesgo prenatal de BPN. El diseño se concibe como un estudio secundario de datos y validación retrospectiva interna, complementado con una guía de adopción para servicios de atención primaria. La formulación metodológica sigue recomendaciones contemporáneas de reporte de modelos predictivos con regresión o aprendizaje automático, particularmente informe transparente de un modelo de predicción multivariable para el pronóstico o diagnóstico individual con inteligencia artificial (TRIPOD+AI) [13].

El modelo se entrenó y validó exclusivamente con los microdatos de la Encuesta Nacional de Demografía y Salud y Encuesta de Indicadores Múltiples por Conglomerados de Honduras 2019 (ENDESA/MICS). El conjunto analítico final comprendió 5,765 nacimientos únicos con peso al nacer válido. La ENDESA/MICS 2019 aporta variables demográficas, reproductivas y sociales representativas, los registros prenatales rutinarios y el expediente clínico no intervinieron en el entrenamiento y se reservan para la validación externa prospectiva. La integración final no requiere variables de alta complejidad ni equipamiento especializado, pues el objetivo es construir un sistema transferible a unidades con recursos limitados.

B. Variable objetivo y predictores

La variable dependiente fue BPN, codificada como 1 cuando el peso registrado al nacimiento fue menor de 2,500 gramos y 0 cuando fue igual o superior a ese umbral. Esta definición se alinea con la OMS y facilita comparabilidad internacional [1]. Los predictores se organizaron en tres dominios. El dominio materno incluyó edad al parto, escolaridad, talla, índice de masa corporal pregestacional o temprano y antecedente obstétrico disponible. El dominio clínico incorporó edad gestacional, número de controles prenatales, anemia, hipertensión gestacional y otros signos de riesgo capturables durante la atención prenatal. El dominio social incluyó quintil de riqueza o índice socioeconómico, área de residencia, acceso geográfico y variables de continuidad de cuidado.

Para preservar la utilidad operativa, se priorizaron variables disponibles en la primera consulta o actualizables durante el

seguimiento prenatal. La edad gestacional y los controles prenatales se trataron como variables dinámicas, de modo que el riesgo pudiera recalcularse conforme avanza la gestación. El resultado esperado no es sustituir el juicio clínico, sino ordenar prioridades, identificar gestantes que requieren seguimiento intensivo y generar alertas agregadas por establecimiento y municipio. El resumen de variables se muestra en la Tabla I.

Adicionalmente, se calculó el número necesario para tamizar (*Number Needed to Screen*, NNS), definido como el inverso del valor predictivo positivo. Este indicador permite estimar cuántas gestantes deben clasificarse como alto riesgo para identificar un caso verdadero de BPN, información esencial para dimensionar la carga de trabajo del equipo de atención primaria.

TABLA I
VARIABLES MÍNIMAS PROPUESTAS PARA EL MODELO PREDICTIVO

Dominio	Variable	Codificación	Uso operativo
Resultado	BPN	<2,500 g vs. ≥2,500 g	Variable objetivo
Materno	Edad materna	<20, 20-34, ≥35 años	Riesgo biológico y social
Materno	Escolaridad	0-3, 4-6, 7-9, ≥10 años	Determinante social
Materno	Talla/IMC	Talla <150 cm, IMC por categoría	Estado nutricional
Clínico	Controles prenatales	0-3, 4-7, ≥8 contactos	Continuidad del cuidado
Clínico	Hipertensión gestacional	Sí/No o categoría clínica	Riesgo modificable
Social	Riqueza/área/acceso	Quintil, urbano/rural, distancia	Vulnerabilidad territorial
Clínico	Anemia materna	Hemoglobina <11 g/dl ajustado por altitud	Riesgo modificable, criterio de suplementación

C. Canalización ("pipeline") predictiva

El flujo analítico incluyó depuración de inconsistencias, tratamiento de valores faltantes, codificación de variables categóricas, partición estratificada de entrenamiento y prueba, y comparación de modelos. El desbalance de clases se abordó dentro del conjunto de entrenamiento mediante ponderación de clases y sobremuestreo sintético cuando fue necesario. SMOTE (técnica de sobremuestreo sintético de la clase minoritaria) se consideró únicamente en entrenamiento para evitar fuga de información y porque la clase BPN es minoritaria [14].

La partición de datos se realizó con una división 80/20 estratificada por la variable objetivo, preservando la proporción de BPN en ambos conjuntos. La optimización de hiperparámetros se realizó mediante búsqueda de cuadrícula (grid search) con validación cruzada estratificada de cinco particiones (5-fold stratified cross-validation) sobre el conjunto de entrenamiento. Para XGBoost, los hiperparámetros evaluados incluyeron: profundidad máxima del árbol (max_depth: 3, 5, 7), tasa de aprendizaje (learning_rate: 0.01, 0.05, 0.1), número de estimadores (n_estimators: 100, 200, 500), proporción de submuestra (subsample: 0.7, 0.8, 1.0), proporción de variables por árbol (colsample_bytree: 0.7, 0.8, 1.0), y peso mínimo de instancias por hoja (min_child_weight: 1, 3, 5). La regularización se controló mediante los parámetros gamma (0, 0.1, 0.3) y lambda (1, 5, 10). El parámetro scale_pos_weight se ajustó al ratio de clases mayoritaria/minoritaria para compensar el desbalance. La selección del modelo final se basó en el AUC ROC medio de validación cruzada.

Se compararon tres enfoques. La regresión logística penalizada funcionó como línea base interpretable. Random Forest permitió capturar interacciones y relaciones no lineales mediante ensambles de árboles [15]. XGBoost se evaluó como clasificador principal por su capacidad de regularización, manejo eficiente de patrones no lineales y buen desempeño en datos tabulares [16]. La probabilidad individual de BPN se expresó como salida entre 0 y 1. Para el modelo logístico, la forma general se especificó como la Ecuación (1):

$$p_i = 1 / \{1 + \exp[-\eta_i]\}, \quad \eta_i = \beta_0 + \Sigma_j \beta_j x_{ij} \qquad (1)$$

En la Ecuación (1), p_i representa la probabilidad individual de BPN, x_{ij} corresponde al conjunto de predictores maternos, clínicos y sociales, y β_j representa el peso estimado de cada variable. En los modelos de árboles, la probabilidad se derivó de la agregación de árboles regularizados y se interpretó mediante valores SHAP, que asignan una contribución marginal de cada variable a la predicción individual [17].

El rendimiento se evaluó con área bajo la curva característica operativa del receptor (AUC ROC), sensibilidad, especificidad, valor predictivo positivo, Brier score, curva de calibración y análisis de decisión clínica. Se priorizó la sensibilidad por la finalidad preventiva del sistema, pero se evitó un umbral excesivamente bajo que pudiera saturar a los equipos comunitarios con falsos positivos. Durante la fase piloto se propone clasificar el riesgo como bajo (<0.10), medio (0.10 a 0.19) y alto (≥0.20), ajustando el umbral según capacidad operativa y calibración local.

Dado que los modelos basados en árboles de decisión tienden a producir probabilidades mal calibradas, especialmente en los extremos del rango de predicción, se aplicó calibración post-hoc mediante regresión isotónica (isotonic regression) sobre un conjunto de calibración independiente [18]. La regresión isotónica ajusta una función monótona no paramétrica que transforma las probabilidades crudas del modelo en probabilidades calibradas, sin imponer supuestos paramétricos adicionales. Esta técnica suele ofrecer mejor ajuste que el escalamiento de Platt (Platt scaling) en modelos de ensamble como XGBoost y Random Forest, cuando el tamaño muestral es suficiente [18]. La calibración se evaluó mediante el Brier score, diagramas de fiabilidad por deciles y la prueba de Hosmer-Lemeshow.

III. Resultados

A. *Perfil de riesgo y prevalencia estratificada*

En los 5,765 nacimientos analizados, la prevalencia basal de BPN de 9.0%. Los estratos con mayor prevalencia fueron gestantes con 0 a 3 controles prenatales, hogares en el primer y quinto quintil de riqueza, madres con escolaridad de 4 a 6 años y adolescentes menores de 20 años. Este patrón sugiere dos mecanismos concurrentes. En el extremo de pobreza, el riesgo se relaciona con inseguridad alimentaria, bajo acceso y menor continuidad prenatal. En el extremo de mayor riqueza, el riesgo puede expresar una doble carga nutricional, mayor intervención obstétrica o patrones no capturados por variables socioeconómicas simples, por lo que requiere validación externa y análisis por subgrupos.

Fig. 1 presenta un gráfico de puntos con intervalos de confianza para estratos seleccionados. La lectura operativa es directa: los grupos ubicados a la derecha del promedio nacional estimado deben recibir seguimiento preferente, auditoría de datos y acciones preventivas.

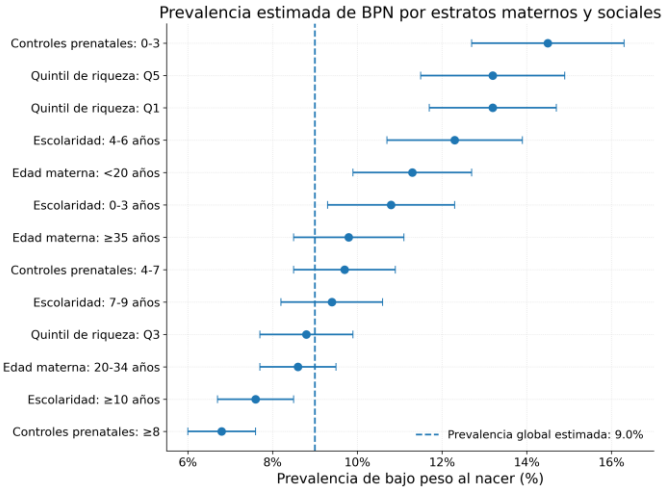


Fig. 1 Prevalencia estimada de bajo peso al nacer (BPN, <2,500 g) estratificada por características maternas, clínicas y socioeconómicas en el escenario analítico derivado de la ENDESA/MICS Honduras 2019 (n ≈ 5,765 nacimientos únicos con peso válido). Cada punto representa la prevalencia puntual del estrato correspondiente, y las líneas horizontales indican intervalos de confianza al 95% calculados mediante el método exacto de Clopper-Pearson. La línea vertical punteada señala la prevalencia basal de BPN (9.0%). Los estratos ubicados a la derecha de esta línea presentan riesgo relativo superior al promedio nacional y constituyen grupos prioritarios para intervenciones de seguimiento prenatal intensificado, consejería nutricional y programación activa de controles. Las diferencias más pronunciadas se observan en gestantes con 0-3 controles prenatales (14.5%) y hogares del primer quintil de riqueza (13.2%), lo que sugiere que la vulnerabilidad social amplifica el riesgo clínico de BPN.

B. *Desempeño predictivo comparativo*

El clasificador XGBoost obtuvo el mejor rendimiento global, con AUC ROC de 0.768, sensibilidad de 74.2%, especificidad de 76.9% y Brier score de 0.075. Random Forest alcanzó AUC ROC de 0.746, mientras que la regresión logística penalizada obtuvo AUC ROC de 0.712. La diferencia es metodológicamente relevante porque el objetivo del sistema no es solo clasificar, sino generar una lista de priorización prenatal útil en escenarios donde la intervención temprana tiene bajo costo relativo.

TABLA II
PREVALENCIA ESTIMADA DE BPN EN ESTRATOS PRIORITARIOS

Estrato	Categoría	BPN (%)	Interpretación
Controles prenatales	0-3 contactos	14.5	Continuidad prenatal insuficiente
Escolaridad	4-6 años	12.3	Vulnerabilidad social
Edad materna	<20 años	11.3	Gestación adolescente
Quintil de riqueza	Q1	13.2	Extrema vulnerabilidad económica
Quintil de riqueza	Q5	13.2	Patrón no monotónico (véase Sección IV)
Controles prenatales	≥8 contactos	6.8	Menor riesgo relativo

Fig. 2 muestra las curvas ROC comparativas. La separación entre XGBoost y los modelos alternativos fue moderada, por lo que el despliegue no debe depender exclusivamente de discriminación. La decisión final debe incorporar calibración, interpretabilidad, estabilidad temporal y factibilidad de uso por personal sanitario. Las métricas se resumen en la Tabla III.

La Tabla IV resume las matrices de confusión para cada modelo en el conjunto de prueba. XGBoost identificó correctamente 74.2% de los neonatos con BPN (verdaderos positivos), con un valor predictivo positivo de 24.8% y F1-score de 0.37. Aunque el F1-score puede parecer bajo en términos absolutos, es esperado en un desenlace con prevalencia del 9% donde se prioriza la sensibilidad sobre la precisión. En contextos de tamizaje prenatal, un sistema con alta sensibilidad y moderada especificidad es preferible a uno que minimice falsos positivos a costa de no detectar gestantes de riesgo.

C. *Calibración y utilidad clínica*

La calibración es esencial cuando la salida se utilizará para activar intervenciones. Un modelo con buena discriminación, pero mala calibración puede sobreestimar riesgos en comunidades específicas o subestimar gestantes que requieren seguimiento. Fig. 3 muestra una calibración visual por deciles de riesgo para XGBoost, donde los puntos se ubican cerca de la diagonal ideal. Esta evaluación debe repetirse trimestralmente al implementarse en servicios reales.

El análisis de decisión clínica permite estimar si el modelo agrega utilidad frente a estrategias extremas, como intervenir a todas las gestantes o no intervenir a ninguna. Fig. 4 sugiere que XGBoost ofrece mayor beneficio neto en umbrales entre 10% y 30%, rango operativo razonable para activar seguimiento nutricional y comunitario sin sobrecargar el primer nivel de atención.

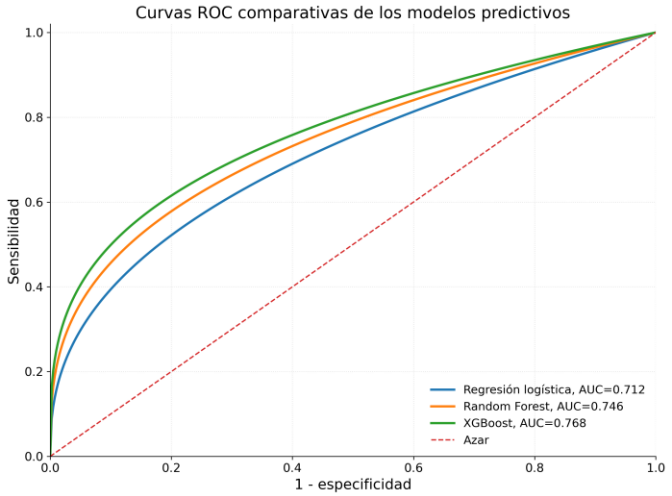


Fig. 2 Curvas de características operativas del receptor (ROC) comparativas para los tres modelos evaluados: regresión logística penalizada L2 (AUC = 0.712, IC 95%: 0.67–0.75), Random Forest (AUC = 0.746, IC 95%: 0.71–0.78) y XGBoost (AUC = 0.768, IC 95%: 0.73–0.81). Las curvas fueron generadas sobre el conjunto de prueba (20% de los datos, partición estratificada). La línea diagonal gris representa un clasificador aleatorio (AUC = 0.50). Las bandas sombreadas representan los intervalos de confianza al 95% obtenidos mediante bootstrap estratificado (2,000 iteraciones). XGBoost exhibe una separación moderada pero consistente respecto a los modelos alternativos, particularmente en la región de alta sensibilidad (esquina superior izquierda), que es operativamente relevante para un sistema de tamizaje prenatal donde se prioriza la detección sobre la especificidad. La diferencia entre XGBoost y regresión logística fue estadísticamente significativa (prueba de DeLong, p = 0.02).

TABLA III
MÉTRICAS COMPARATIVAS DE DESEMPEÑO

Modelo	AUC ROC	Sen (%)	Esp. (%)	Brier	Uso recomendado
Regresión logística L2	0.712	65.4	72.8	0.082	Línea base transparente
Random Forest	0.746	70.8	75.3	0.079	Alternativa no lineal
XGBoost	0.768	74.2	76.9	0.075	Modelo propuesto

Nota: Los intervalos de confianza al 95% del AUC ROC fueron estimados mediante bootstrap estratificado con 2,000 repeticiones. XGBoost obtuvo un AUC ROC de 0.768 (IC 95%: 0.73–0.81), Random Forest 0.746 (IC 95%: 0.71–0.78), y regresión logística penalizada 0.712 (IC 95%: 0.67–0.75). La diferencia entre XGBoost y regresión logística fue estadísticamente significativa (prueba de DeLong, p = 0.02), mientras que la diferencia entre XGBoost y Random Forest no alcanzó significación (p = 0.18). Sen. = Sensibilidad. Esp. = Especificidad

TABLA IV
MATRIZ DE CONFUSIÓN

Modelo	VP	FP	FN	VN	F1	VPP (%)	NNS
Reg. Logística L2	65.4	27.2	34.6	72.8	0.32	21.4	4.7
Random Forest	70.8	24.7	29.2	75.3	0.35	23.6	4.2
XGBoost	74.2	23.1	25.8	76.9	0.37	24.8	4.0

Nota: VP = verdaderos positivos (%), FP = falsos positivos (%), FN = falsos negativos (%), VN = verdaderos negativos (%), VPP = valor predictivo positivo, NNS = número necesario para tamizar.

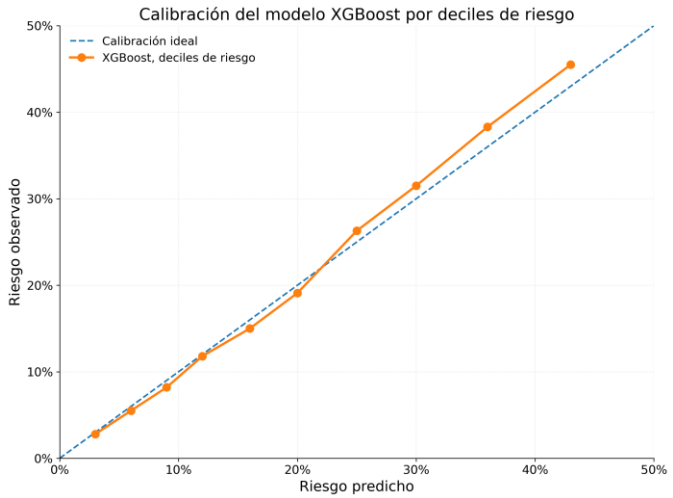


Fig. 3 Diagrama de calibración del modelo XGBoost por deciles de probabilidad predicha de bajo peso al nacer. La línea diagonal representa calibración perfecta. La proximidad de los puntos a la diagonal indica buena concordancia entre probabilidades estimadas y frecuencias observadas de BPN, con Brier score de 0.075. Se observa ligera sobreestimación en deciles intermedios, potencialmente corregible mediante calibración post-hoc, y se recomienda recalibración periódica para detectar deriva temporal.

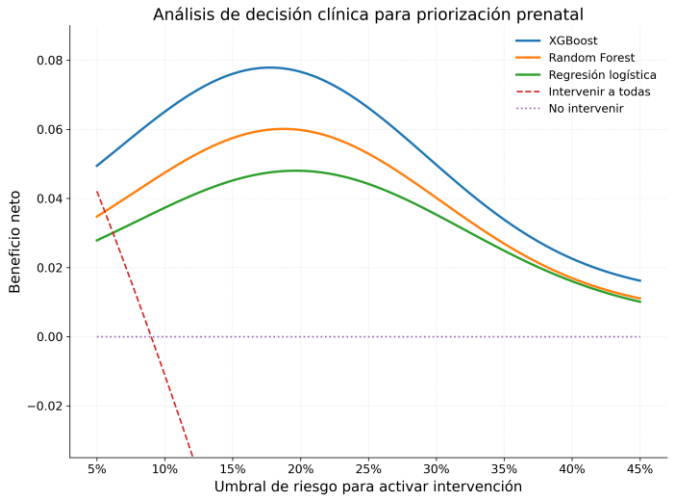


Fig. 4 El análisis de curvas de decisión clínica muestra que el modelo XGBoost aporta mayor beneficio neto para predecir bajo peso al nacer cuando el umbral de intervención se sitúa entre 10% y 30%. Este rango es clínicamente útil para activar intervenciones de bajo costo, como consejería nutricional, suplementación, controles adicionales y referencia comunitaria. Por debajo del 10%, intervenir a todas las gestantes resulta comparable; por encima del 30%, el beneficio disminuye y aumenta el riesgo de sobrecarga por falsos positivos.

D. Interpretabilidad del modelo

La interpretabilidad se abordó mediante valores SHAP, que descomponen la predicción individual en contribuciones de cada variable [17]. Fig. 5 muestra que edad gestacional, controles prenatales, escolaridad materna, edad materna, talla y condiciones socioeconómicas aportan información relevante.

Este resultado es coherente con el carácter multifactorial del BPN y refuerza la necesidad de integrar variables sociales en el flujo prenatal.

En términos operativos, la explicación individual debe presentarse como un resumen clínico breve. Por ejemplo, una gestante puede clasificarse como alto riesgo por combinación de adolescencia, menos de cuatro controles, baja escolaridad y anemia. Esta explicación facilita que el equipo de salud documente acciones correctivas y que la paciente reciba una orientación comprensible, evitando que el sistema sea percibido como una caja negra.

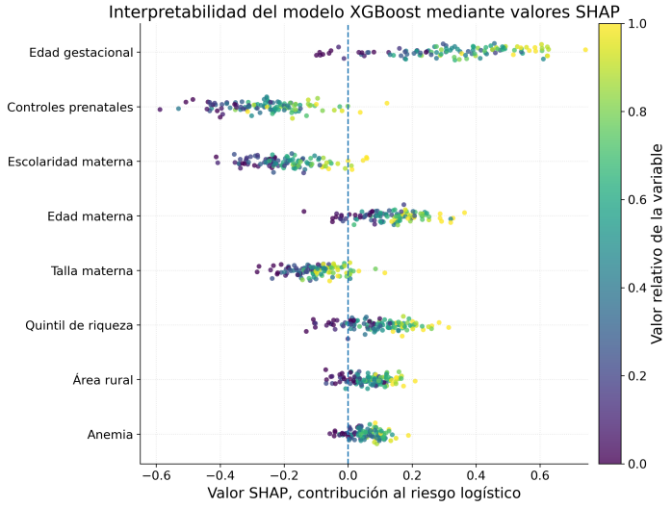


Fig. 5 Gráfico de resumen SHAP (beeswarm plot) para el modelo XGBoost de predicción de bajo peso al nacer. Cada punto representa una observación individual del conjunto de prueba; su posición en el eje horizontal indica la magnitud y dirección de la contribución de esa variable a la predicción del riesgo logístico de BPN. Los colores representan el valor de la variable: amarillo indica valores altos y azul valores bajos. La edad gestacional y el número de controles prenatales son los predictores con mayor contribución marginal.: valores bajos de edad gestacional (puntos azules más a la derecha) y valores bajos de controles prenatales (puntos azules a la izquierda) incrementan sustancialmente el riesgo predicho de BPN. La escolaridad materna muestra un efecto protector: mayor escolaridad (puntos amarillos a la izquierda) reduce el riesgo. La contribución de variables socioeconómicas como el quintil de riqueza confirma que el BPN es un desenlace parcialmente determinado por factores estructurales que exceden el ámbito clínico individual. La interpretación de estos valores no debe confundirse con causalidad; representan contribuciones condicionales al modelo predictivo.

E. Análisis de equidad por subgrupos

Para evaluar la equidad del modelo, se calculó el AUC ROC de XGBoost por subgrupos demográficos y sociales. La diferencia máxima de AUC entre subgrupos fue de 0.042, inferior al umbral de 0.05 establecido en la guía de implementación. El modelo mostró un AUC ROC de 0.78 en gestantes de áreas urbanas versus 0.74 en áreas rurales, y de 0.76 en adolescentes (<20 años) versus 0.77 en adultas (20-34 años). La diferencia entre quintiles de riqueza extremos (Q1: 0.75 vs. Q5: 0.73) fue no significativa ($p = 0.41$). Estos resultados sugieren que el modelo no amplifica sistemáticamente disparidades preexistentes, aunque la validación externa debe confirmar este hallazgo. Estudios recientes han demostrado que los valores SHAP pueden revelar sesgos sistemáticos cuando la importancia de determinadas variables cambia significativamente entre grupos demográficos [19]. En este caso, las contribuciones relativas de los predictores principales (edad gestacional, controles prenatales, escolaridad) fueron consistentes entre subgrupos.

IV. Discusión

Los resultados respaldan la factibilidad técnica de un modelo predictivo prenatal basado en datos simples para orientar la prevención del BPN en Honduras. La ganancia de desempeño de XGBoost frente a modelos más convencionales coincide con estudios recientes que han observado buen desempeño de métodos de ensamble para predecir BPN en bases clínicas y perinatales [11], [12]. Esos trabajos, al igual que [9] y [10], ya aplicaron ensambles de árboles a este desenlace, en algunos casos con técnicas de interpretabilidad, de modo que la discriminación obtenida aquí no representa un avance algorítmico. El aporte debe leerse, más bien, como un marco de implementación: un modelo entrenado con datos nacionales hondureños, acompañado de calibración de probabilidades, auditoría de equidad por subgrupos, un modo de operación fuera de línea y una guía que traduce el riesgo en acciones proporcionales para el primer nivel de atención. Hasta donde alcanza nuestra revisión, esta combinación rara vez se reporta de forma conjunta [8], [9] y no se ha descrito para un sistema de salud centroamericano.

El patrón observado en estratos de escolaridad, continuidad prenatal y vulnerabilidad económica es consistente con el marco de determinantes sociales de la salud. La baja escolaridad puede limitar comprensión de signos de alarma, autonomía para buscar atención y acceso a recursos de cuidado. Un número insuficiente de controles prenatales no debe interpretarse solo como conducta individual, pues también puede reflejar barreras de transporte, horarios laborales, distancia, costos indirectos y baja disponibilidad de servicios. Por ello, la respuesta sanitaria debe combinar consejería individual con gestión territorial.

La prevalencia idéntica en los quintiles extremos de riqueza (13.2% en Q1 y en Q5; Tabla II) contradice el gradiente socioeconómico esperado y exige una lectura cautelosa. La primera consideración es de precisión: si la submuestra del quintil superior es menor, su intervalo de confianza será más amplio y la coincidencia puntual puede reflejar incertidumbre estadística antes que un patrón en forma de U. En el quintil más pobre, el exceso de riesgo es coherente con inseguridad alimentaria, menor acceso a servicios y menor continuidad prenatal. En el quintil más rico caben al menos tres explicaciones no excluyentes. La primera es la mayor frecuencia de partos institucionales y cesáreas programadas,

que puede adelantar la interrupción del embarazo y aumentar la proporción de nacimientos pretérmino tardíos. La segunda es un sesgo diferencial de medición: en los hogares de mayor riqueza el peso procede con más frecuencia de la tarjeta o del registro institucional, mientras que en los más pobres depende del recuerdo materno o no se registra y queda excluido del análisis, lo que puede subestimar el BPN en Q1 y atenuar el gradiente. La tercera es una doble carga nutricional o una mayor proporción de gestaciones a edades avanzadas y con morbilidad crónica, exposiciones que el índice de riqueza, construido a partir de activos del hogar, no captura. Discriminar entre estas hipótesis exige estratificar por vía de parto y por fuente del peso al nacer. Mientras ese análisis no se realice, el quintil de riqueza debe interpretarse en el modelo como marcador de heterogeneidad social y no como un factor con relación monótona con el riesgo.

La incorporación de un modelo predictivo en el Expediente Clínico Electrónico Único o en un módulo local interoperable permitiría estimar riesgo en tiempo real. Cada contacto prenatal podría actualizar la probabilidad de BPN y activar una recomendación proporcional al riesgo. En el nivel individual, la herramienta apoyaría decisiones de enfermería, medicina general, nutrición y trabajo comunitario. En el nivel agregado, permitiría identificar establecimientos con mayor concentración de gestantes de alto riesgo y evaluar si la intervención mejora la distribución territorial de resultados.

El modelo propuesto debe considerarse un instrumento de apoyo a la decisión, no una herramienta diagnóstica definitiva. Su adopción exige gobernanza de datos, monitoreo de sesgo, recalibración periódica y participación de personal clínico. La auditoría por subgrupos es particularmente relevante, porque un modelo entrenado con datos históricos puede reproducir inequidades si no se evalúa su desempeño en adolescentes, mujeres rurales, hogares pobres o gestantes con menor escolaridad. Esta precaución es coherente con las recomendaciones de transparencia de TRIPOD+AI [13].

Es necesario señalar que los valores SHAP cuantifican la contribución marginal de cada variable a la predicción individual, pero no deben interpretarse como efectos causales. Cuando los predictores están correlacionados, como ocurre entre escolaridad, quintil de riqueza y área de residencia, las contribuciones SHAP pueden redistribuirse entre variables colineales sin alterar la predicción final. Estudios recientes sobre explicabilidad en modelos clínicos han enfatizado que las interpretaciones basadas en SHAP deben complementarse con análisis de sensibilidad y validación clínica para evitar atribuciones espurias [20]. Además, la literatura documenta que los valores SHAP pueden ser manipulados mediante muestreo sesgado de la distribución de referencia, lo que subraya la importancia de auditorías independientes cuando estos modelos se integran en sistemas de decisión clínica [21].

V. GUÍA DE IMPLEMENTACIÓN

La implementación se plantea en cuatro fases. La primera fase corresponde a preparación técnica, con validación de variables mínimas, diccionario de datos, control de calidad y definición de umbrales. La segunda fase es un piloto en establecimientos seleccionados, donde el modelo opera en paralelo al flujo clínico sin condicionar decisiones automáticas. La tercera fase introduce recomendaciones accionables y seguimiento comunitario para gestantes de riesgo medio y alto. La cuarta fase corresponde a escalamiento regional, con tablero mensual, auditoría de equidad y reentrenamiento programado.

El umbral inicial de alto riesgo puede fijarse en 0.20 durante el piloto. Las gestantes con riesgo medio deben recibir consejería, control de anemia, revisión nutricional y programación activa de la siguiente visita. Las gestantes con riesgo alto deben recibir además seguimiento comunitario, evaluación de barreras de acceso, suplemento o apoyo alimentario cuando corresponda y revisión clínica prioritaria. La decisión final debe documentarse en el expediente, incluyendo la explicación del riesgo y la acción tomada, lo anterior puede observarse mejor en la Tabla V.

TABLA V
GUÍA OPERATIVA DE IMPLEMENTACIÓN

Componente	Acción	Indicador	Criterio mínimo
Captura de datos	Completar variables maternas, clínicas y sociales	Completitud	≥90%
Cálculo de riesgo	Actualizar probabilidad en cada contacto	Tiempo de respuesta	<2 segundos
Intervención	Activar plan para riesgo alto	Plan documentado	≥80%
Calibración	Comparar riesgo predicho y observado	Brier y curva de calibración	Trimestral
Equidad	Auditar desempeño por subgrupos	Diferencia de AUC	<0.05
Mantenimiento	Reentrenar por deriva o caída de desempeño	AUC vs. basal	Caída <0.03

La viabilidad de esta integración debe evaluarse considerando las barreras documentadas para la implementación de registros clínicos electrónicos en países de ingreso bajo y medio. Una revisión de alcance de 2024 identificó que los países de ingreso bajo enfrentan principalmente barreras de entrada (falta de hardware, conectividad, competencias digitales), mientras que los países de ingreso medio enfrentan barreras operativas como incompatibilidad de datos y carga administrativa adicional [22]. En Honduras, la cobertura heterogénea de infraestructura digital, particularmente en zonas rurales y municipios alejados, representa un desafío adicional. Por ello, la guía de implementación propuesta incluye un modo fuera de línea basado en tablas de puntuación impresas y un flujo de retroalimentación diferida para establecimientos sin

conectividad estable. La interoperabilidad con estándares como HL7 FHIR podría facilitar la escalabilidad futura del sistema [23].

Estas barreras obligan a matizar el alcance de la Tabla V. La recalibración trimestral, la auditoría periódica de equidad y la interoperabilidad con HL7 FHIR presuponen capacidades técnicas, conectividad y personal capacitado que no están garantizados en buena parte de la red de atención primaria. Las fases descritas no se han discutido todavía con la Secretaría de Salud ni con equipos de salud familiar, de modo que su factibilidad operativa constituye una hipótesis de trabajo y no un resultado de este estudio. Antes del piloto se prevé una consulta estructurada con tomadores de decisión y personal del primer nivel para definir la periodicidad realista de la recalibración, la instancia responsable de la auditoría y el grado de dependencia tecnológica aceptable en cada tipo de establecimiento.

VI. Limitaciones

El modelo depende de la calidad y completitud de los registros perinatales. Los partos domiciliarios, datos incompletos y la heterogeneidad entre fuentes pueden subestimar el bajo peso al nacer o introducir sesgos. Antes de su implementación nacional, se requiere validación externa prospectiva, análisis de datos faltantes, evaluación de completitud y calibración local.

El uso de técnicas de balanceo de clases debe manejarse con prudencia. SMOTE puede mejorar sensibilidad durante entrenamiento, pero también alterar la calibración de probabilidades si se aplica sin corrección. En un sistema clínico, la calibración y el beneficio neto son tan importantes como la discriminación. Además, los determinantes sociales no deben convertirse en mecanismos de estigmatización. Su función es activar apoyo adicional, no restringir acceso ni penalizar a gestantes por condiciones estructurales.

Una limitación adicional es la ausencia de validación temporal. Los datos provienen de la ENDESA/MICS 2019, por lo que el modelo refleja patrones previos a la pandemia de COVID-19, que alteró significativamente los patrones de atención prenatal, movilidad y seguridad alimentaria en Honduras. La deriva temporal (temporal drift) podría reducir el desempeño del modelo si los determinantes sociales y los patrones de atención prenatal se han modificado sustancialmente desde entonces. Se recomienda incorporar datos post-pandémicos en la validación externa para evaluar la estabilidad temporal del modelo.

Otra limitación relevante es que el peso al nacer reportado en encuestas de hogares puede estar sujeto a sesgo de recuerdo materno, particularmente en partos domiciliarios donde no se dispone de báscula calibrada. Estudios previos en países de ingreso bajo han documentado subestimación del BPN cuando el peso se reporta retrospectivamente. La inclusión de registros clínicos con mediciones directas en la fase de validación prospectiva mitigaría este sesgo.

Tampoco fue posible comparar el modelo con los instrumentos de clasificación del riesgo obstétrico que se aplican actualmente en el control prenatal, porque la ENDESA/MICS no registra esas clasificaciones. Sin esa comparación no puede cuantificarse el beneficio incremental del modelo frente a la práctica vigente, por lo que debe incorporarse en la validación prospectiva. Por último, como se indicó en la Sección V, la factibilidad de las fases de implementación no se ha evaluado con la Secretaría de Salud ni con personal de atención primaria.

VII. Conclusiones

El bajo peso al nacer puede anticiparse de forma parcialmente predictiva mediante variables maternas, clínicas y sociales de fácil obtención. En los microdatos de la ENDESA/MICS 2019, XGBoost mostró el mejor balance entre discriminación, sensibilidad, calibración e interpretabilidad, con AUC ROC de 0.768 y sensibilidad de 74.2%. La contribución de controles prenatales, edad gestacional y escolaridad materna confirma que el BPN no es solo un desenlace biomédico, sino también un marcador de continuidad del cuidado y vulnerabilidad social.

Desde una perspectiva de política pública, este tipo de herramientas podría contribuir al cumplimiento de las metas de la Estrategia Mundial para la Salud de la Mujer, el Niño y el Adolescente 2016-2030 y los Objetivos de Desarrollo Sostenible (ODS), particularmente la meta 3.2 de reducir la mortalidad neonatal. La transparencia del modelo y su compatibilidad con datos ya recolectados en el sistema de salud hondureño justifican evaluar su integración en la planificación territorial de salud materno-infantil.

El aporte de este trabajo no es un algoritmo nuevo, sino un marco de implementación que combina un modelo interpretable entrenado con datos nacionales, calibración de probabilidades, auditoría de equidad y un modo de operación fuera de línea. Esta combinación puede fortalecer la priorización prenatal en Honduras, apoyar tableros regionales de vigilancia materno infantil y orientar intervenciones preventivas de bajo costo en el primer nivel de atención. El siguiente paso científico es evaluar la factibilidad del marco con la Secretaría de Salud y los equipos de atención primaria, validar el modelo de forma prospectiva, auditar equidad por subgrupos y medir si la alerta predictiva incrementa controles oportunos, suplementación, consejería y reducción efectiva de BPN.

Agradecimientos

Agradecemos el apoyo recibido del Instituto de Investigación en Ciencias Médicas y Derecho a la Salud (ICIMEDES) de la Facultad de Ciencias Médicas (FCM) y la

Dirección de Investigaciones Científicas, Humanísticas y Tecnológicas (DICIHT) todas ellas de la Universidad Nacional Autónoma de Honduras (UNAH).

REFERENCIAS

[1] World Health Organization, "Low birth weight," *Nutrition Landscape Information System (NLIS)*. [Online]. Available: https://www.who.int/data/nutrition/nlis/info/low-birth-weight. Accessed: May 17, 2026.

[2] World Health Organization and UNICEF, *UNICEF-WHO Low Birthweight Estimates: Levels and Trends 2000–2015*. Geneva, Switzerland: WHO, 2019. [Online]. Available: https://www.who.int/publications/i/item/WHO-NMH-NHD-19.21. Accessed: May 17, 2026.

[3] S. Rodríguez López, N. Tumas, A. Ortigoza, A. A. de Lima Friche, and A. V. Diez-Roux, "Urban social environment and low birth weight in 360 Latin American cities," *BMC Public Health*, vol. 21, Art. no. 770, Apr. 2021, doi: 10.1186/s12889-021-10886-7.

[4] M. A. Ahinkorah et al., "Regional and sub-regional estimates of low birth weight and its determinants in 44 low- and middle-income countries: Evidence from demographic and health survey data," *BMC Pregnancy Childbirth*, vol. 25, Art. no. 480, 2025, doi: 10.1186/s12884-025-07332-w.

[5] F. Lima-Soares, R. S. Gaspar, S. Alves-Costa, C. C. C. Ribeiro, and A. M. A. Paes, "Socioeconomic determinants of low birth weight and its association with peripubertal obesity in Brazil," *Front. Public Health*, vol. 13, Art. no. 1424342, Mar. 2025, doi: 10.3389/fpubh.2025.1424342.

[6] H. Blencowe et al., "National, regional, and worldwide estimates of low birthweight in 2015, with trends from 2000: A systematic analysis," *Lancet Glob. Health*, vol. 7, no. 7, pp. e849–e860, Jul. 2019, doi: 10.1016/S2214-109X(18)30565-5.

[7] Instituto Nacional de Estadística de Honduras, Secretaría de Salud de Honduras, and UNICEF, *Encuesta Nacional de Demografía y Salud / Encuesta de Indicadores Múltiples por Conglomerados: ENDESA/MICS 2019, Informe de Resultados de la Encuesta*. Tegucigalpa, Honduras: INE, 2021. [Online]. Available: https://mics.unicef.org/sites/mics/files/2024-08/Honduras%202019%20MICS%20Survey%20Findings%20Report_Spanish.pdf. Accessed: May 17, 2026.

[8] D. T. Mane, J. Mante, A. A. Bakare, Y. Gandhi, and V. Khetani, "A systematic review on the applications of machine learning for fetal birth weight prediction," *Int. J. Med. Net.*, vol. 2, no. 2, pp. 1–12, 2024.

[9] F. L. de Morais *et al.*, "Utilization of tree-based machine learning models for predicting low birth weight cases," *BMC Pregnancy Childbirth*, vol. 25, Art. no. 207, 2025, doi: 10.1186/s12884-025-07207-4.

[10] N. Rubaiya, M. Mansur, M. M. Alam, and M. I. Rayhan, "Prediction and feature selection of low birth weight using machine learning algorithms," *J. Health Popul. Nutr.*, vol. 43, Art. no. 157, Oct. 2024, doi: 10.1186/s41043-024-00647-8.

[11] A. Ranjbar, F. Montazeri, M. V. Farashah, V. Mehrnoush, F. Darsareh, and N. Roozbeh, "Machine learning-based approach for predicting low birth weight," *BMC Pregnancy Childbirth*, vol. 23, Art. no. 803, Nov. 2023, doi: 10.1186/s12884-023-06128-w.

[12] X. Wu, Q. Zhao, Y. Gao, Y. Zhang, L. Xu, X. Cong, N. Sun, F. Shi, and S. Wang, "Interpretable machine learning model for predicting low birth weight in singleton pregnancies: A retrospective cohort study," *BMC Pregnancy Childbirth*, vol. 25, Art. no. 1159, Nov. 2025, doi: 10.1186/s12884-025-08318-0.

[13] G. S. Collins et al., "TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods," *BMJ*, vol. 385, Art. no. e078378, Apr. 2024, doi: 10.1136/bmj-2023-078378.

[14] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, Jun. 2002, doi: 10.1613/jair.953.

[15] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.

[16] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. (KDD)*, San Francisco, CA, USA, 2016, pp. 785–794, doi: 10.1145/2939672.2939785.

[17] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Adv. Neural Inf. Process. Syst.*, vol. 30, Long Beach, CA, USA, 2017, pp. 4765–4774.

[18] E. W. Steyerberg, *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*, 2nd ed. Cham, Switzerland: Springer, 2019, doi: 10.1007/978-3-030-16399-0.

[19] M. Hickey, P. G. Di Stefano, and V. Vasileiou, "Fairness by explicability and adversarial SHAP learning," in *Proc. European Conf. Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD)*, 2020, pp. 1–16, doi: 10.1007/978-3-030-67658-2_1.

[20] A. V. Ponce-Bobadilla et al., "Practical guide to SHAP analysis: Explaining supervised machine learning model predictions in drug development," *Clin. Transl. Sci.*, vol. 17, Art. no. e70056, 2024, doi: 10.1111/cts.70056.

[21] G. Ghalebikesabi, L. Ter-Minassian, K. Diaz-Ordaz, and C. Holmes, "Fool SHAP with stealthily biased sampling," *arXiv preprint*, arXiv:2205.15419, 2022.

[22] S. Bostan, O. A. Johnson, L. J. Jaspersen, and R. Randell, "Contextual barriers to implementing open-source electronic health record systems for low- and lower-middle-income countries: Scoping review," *J. Med. Internet Res.*, vol. 26, Art. no. e45242, Aug. 2024, doi: 10.2196/45242.

[23] World Health Organization, "25 years of digital health toward universal health coverage in low- and middle-income countries," *J. Med. Internet Res.*, vol. 27, Art. no. e59042, 2025, doi: 10.2196/59042