Explainable Artificial Intelligence to Detect Gaps in the Prenatal Care Continuum: Methodological Proposal with an Emphasis on Honduras

Carlos Agudelo Santos[1]; Gustavo Cruz Martinez[2]; Rossana Carias Escoto[3]
[1,3]Universidad Nacional Autónoma de Honduras, Honduras, *carlos.agudelo@unah.edu.hn, rossana_escoto@unah.edu.hn*
[2]Instituto Hondureño de Seguridad Social, Honduras, *gustavo.cruz@ihss.hn*

Abstract– ***Honduras presents relevant gaps in the prenatal care continuum. According to ENDESA/MICS 2019, 96.1% of women with a live birth in the previous two years received at least one antenatal care contact from skilled health personnel; however, only 65.9% initiated care before three months of gestation, 88.3% completed four or more visits, and 39.0% reached eight or more antenatal contacts. This progressive loss between initial access, timely first contact, and continuity of follow-up reflects a structural adherence problem in prenatal care. This paper proposes a methodological framework based on explainable artificial intelligence to support the future identification of discontinuity risk in antenatal care, integrating predictive models, global and local explainability methods, equity auditing, and human oversight. Because individual-level microdata were unavailable, the pipeline was executed on a synthetic cohort (n=20,000) calibrated to ENDESA/MICS 2019 indicators to demonstrate operational feasibility: penalized logistic regression, Random Forest and XGBoost reached equivalent discrimination (AUC 0.727) with adequate calibration, and SHAP recovered the simulated risk structure. These results are illustrative, not empirical estimates. A pilot validation protocol is specified, including subgroup calibration and equity auditing, participation of health workers and pregnant women, and indicators of impact on care continuity at the first level of care in Honduras.***

Keywords-- ***explainable artificial intelligence, prenatal care, continuum of care, Honduras, Digital maternal health.***

Inteligencia Artificial explicable para detectar brechas en el continuo de atención prenatal: Propuesta metodológica con énfasis en Honduras

Carlos Agudelo Santos[1]; Gustavo Cruz Martinez[2]; Rossana Carias Escoto[3]
[1,3]Universidad Nacional Autónoma de Honduras, Honduras, *carlos.agudelo@unah.edu.hn, rossana_escoto@unah.edu.hn*
[2]Instituto Hondureño de Seguridad Social, Honduras, *gustavo.cruz@ihss.hn*

Resumen–Honduras presenta brechas relevantes en el continuo de atención prenatal. Según ENDESA/MICS 2019, el 96.1% de las mujeres con un hijo nacido vivo en los dos años previos recibió al menos una atención prenatal por personal de salud calificado; sin embargo, solo el 65.9% inició la atención antes de los tres meses de gestación, el 88.3% completó cuatro o más visitas y el 39.0% alcanzó ocho o más contactos prenatales. Esta pérdida progresiva entre captación inicial, oportunidad del primer contacto y continuidad del seguimiento evidencia un problema estructural de adherencia al cuidado prenatal. Se propone un marco metodológico basado en inteligencia artificial explicable para diseñar un sistema futuro de identificación de riesgo de discontinuidad prenatal, integrando modelos predictivos, técnicas de explicabilidad global y local, auditoría de equidad y supervisión humana. Ante la falta de microdatos individuales, el pipeline se ejecutó sobre una cohorte sintética (n=20,000) calibrada con indicadores de ENDESA/MICS 2019 para demostrar su viabilidad operativa: la regresión logística penalizada, Random Forest y XGBoost alcanzaron una discriminación equivalente (AUC 0.727) con calibración adecuada, y SHAP recuperó la estructura de riesgo simulada. Estos resultados son ilustrativos y no estimaciones empíricas. Se especifica además un protocolo de validación piloto con auditoría de calibración y equidad por subgrupos, participación de personal sanitario y gestantes, e indicadores de impacto sobre la continuidad prenatal en el primer nivel de atención de Honduras.

Palabras clave-- inteligencia artificial explicable; atención prenatal; continuo de atención; Honduras; salud materna digital.

I. Introducción

La mortalidad materna continúa siendo un indicador centinela de desigualdad sanitaria y desempeño del sistema de salud. A nivel global, la OMS ha señalado que la mayoría de las muertes maternas son prevenibles cuando existen servicios oportunos, personal calificado, continuidad asistencial y capacidad de respuesta ante complicaciones obstétricas. En Honduras, el problema no se limita al acceso inicial a la atención prenatal, sino a la oportunidad y continuidad del seguimiento durante el embarazo. La Encuesta Nacional De Demografía Y Salud / Encuesta De Indicadores Múltiples Por Conglomerados (ENDESA/MICS 2019) muestra que, aunque la atención prenatal inicial por personal calificado alcanza 96.1%, la cobertura disminuye al considerar el inicio temprano, el número de contactos y la adherencia al estándar ampliado de ocho contactos prenatales [1]-[3].

La mortalidad materna no debe interpretarse únicamente como un desenlace clínico individual, sino como un evento centinela de fallas acumuladas en el sistema de salud, incluyendo barreras de acceso, demoras en la búsqueda de atención, discontinuidad del seguimiento prenatal y limitaciones en la calidad resolutiva de los servicios. La OMS enfatiza que la mayoría de las muertes maternas son prevenibles mediante atención oportuna, personal calificado y sistemas capaces de identificar tempranamente complicaciones durante el embarazo, el parto y el puerperio [4]. En consecuencia, el análisis del continuo de atención prenatal ofrece una vía operativa para anticipar riesgos, priorizar intervenciones territoriales y reducir desigualdades evitables en salud materna.

El continuo de atención prenatal comprende una secuencia de contactos entre la gestante y el sistema de salud que, según las recomendaciones vigentes de la Organización Mundial de la Salud (OMS), debe incluir al menos ocho visitas iniciadas antes de las doce semanas de gestación [5]. Este marco supera el anterior estándar de cuatro visitas (ANC4+) y reconoce que la calidad y la oportunidad de cada contacto son tan determinantes como su número [6].

En Honduras, ENDESA/MICS 2019 documentó que el 96.1% de las mujeres con un hijo nacido vivo en los dos años previos recibió al menos una atención prenatal por personal de salud calificado. Sin embargo, la continuidad del cuidado fue menor: 88.3% completó cuatro o más visitas, 80.6% inició la atención antes de los cuatro meses, 65.9% antes de los tres meses y solo 39.0% alcanzó ocho o más contactos prenatales [3]. Además, la cobertura de ocho o más contactos mostró variación territorial importante, con valores de 20.0% en Gracias a Dios, 20.5% en Olancho y 62.6% en el Distrito Central [3].

Desde una perspectiva analítica, las brechas en el continuo de atención prenatal pueden operacionalizarse en tres transiciones críticas: captación temprana, cumplimiento mínimo de cuatro contactos y adherencia al estándar ampliado de ocho contactos. Esta distinción es metodológicamente relevante porque cada transición puede estar asociada a determinantes distintos: la captación temprana puede reflejar

alfabetización sanitaria, reconocimiento del embarazo y disponibilidad inicial de servicios, mientras que la continuidad de los contactos puede depender de costos indirectos, transporte, trato percibido, condiciones laborales y capacidad de seguimiento comunitario. Estos indicadores sugieren que el principal problema está más allá de la entrada inicial al sistema, se encuentra en la dificultad de la continuidad efectiva y oportuna del cuidado prenatal.

Estas brechas en el continuo no son aleatorias. Responden a patrones estructurales relacionados con la ubicación geográfica, el nivel educativo, la pertenencia a pueblos indígenas y afrodescendientes, la disponibilidad de transporte y la calidad percibida de la atención [2], [3]. Detectar tempranamente a las gestantes con mayor riesgo de abandono del continuo es, por tanto, un problema analítico con consecuencias clínicas directas.

La inteligencia artificial (IA) ha emergido como una herramienta poderosa para la predicción de riesgos en salud materna [7], [8]. Sin embargo, los modelos de aprendizaje automático de alto rendimiento como los bosques aleatorios o el *gradient boosting* presentan el conocido problema de la "caja negra": producen predicciones precisas, pero no explican las razones subyacentes, lo que limita su adopción clínica y genera desconfianza institucional [9], [10]. La inteligencia artificial (IA) explicable (Explainable Artificial Intelligence, XAI) busca resolver esta tensión ofreciendo representaciones comprensibles de las decisiones del modelo para médicos, enfermeras y tomadores de decisión [11], [12].

La utilidad de los modelos predictivos en salud materna depende no solo de su capacidad discriminativa, sino también de su calibración, interpretabilidad y factibilidad de incorporación en flujos reales de atención. En escenarios de salud pública, una alerta predictiva debe permitir identificar qué factores contribuyen al riesgo estimado, diferenciar determinantes modificables de no modificables y orientar acciones concretas, tales como seguimiento comunitario, priorización de citas, derivación temprana o intervención territorial. Métodos de explicación local y global, como LIME y SHAP, permiten traducir modelos complejos en perfiles de riesgo comprensibles para equipos clínicos y gestores sanitarios. No obstante, en decisiones de alto impacto, también debe considerarse el uso de modelos inherentemente interpretables y mecanismos de auditoría, equidad y gobernanza, conforme a las recomendaciones contemporáneas sobre IA responsable en salud [9], [13].

En el contexto centroamericano, la aplicación de XAI en salud materna es incipiente. No existen, hasta donde conocen los autores, modelos XAI implementados en el sistema de información perinatal hondureño. La presente propuesta metodológica llena ese vacío al formalizar un canalización o pipeline analítico completo, desde la captura de datos hasta la generación de alertas clínicas interpretables, orientado específicamente a detectar brechas en el continuo de ANC en Honduras.

II. Materiales y métodos

A. *Tipo de estudio y justificación del diseño*

Se realizó una propuesta metodológica con análisis documental de indicadores publicados, diseño apropiado cuando el objeto de investigación es la especificación de un f*ramework* analítico cuya validación empírica requiere acceso a microdatos no disponibles en el presente trabajo. La propuesta integra cuatro componentes: caracterización cuantitativa de las brechas en el continuo prenatal hondureño usando indicadores publicados, especificación del pipeline de modelado XAI, simulación de su funcionamiento con una cohorte sintética y protocolo de validación y despliegue en el primer nivel de atención.

B. *Fuentes consultadas*

La caracterización de las brechas en el continuo prenatal se basó principalmente en ENDESA/MICS 2019, que reporta indicadores nacionales, urbanos, rurales, departamentales y por características sociodemográficas de atención prenatal [3]. Para el marco normativo se utilizó la recomendación de la OMS sobre atención prenatal para una experiencia positiva del embarazo, que establece un modelo mínimo de ocho contactos y primer contacto durante el primer trimestre [5].

Para el componente de inteligencia artificial explicable se consideraron referencias metodológicas sobre interpretabilidad, explicabilidad post hoc, aprendizaje automático en salud y gobernanza ética de IA [7]–[15]. Para la especificación técnica del *pipeline* se emplearon las referencias originales de Random Forest, XGBoost, SHAP, TreeSHAP y LIME [16]–[20]. La validación futura del modelo se plantea conforme a Información Transparente de un Modelo de Predicción Multivariable para el Pronóstico o Diagnóstico Individual + Inteligencia Artificial (TRIPOD+AI) y Herramienta de Evaluación del Riesgo de Sesgo y Aplicabilidad de Modelos de Predicción + Inteligencia Artificial (PROBAST+AI), dado que estas guías actualizan los criterios de reporte, riesgo de sesgo y aplicabilidad para modelos predictivos basados en regresión, aprendizaje automático o IA [21]-[23]. Ver Tabla I con detalles de uso.

TABLA I
FUENTES DE DATOS ABIERTAS CONSIDERADAS PARA ANÁLISIS FUTURO

Fuente	Uso en el estudio	Nivel de información	Variables relevantes	Estado
ENDESA/MICS 2019 [3]	Caracterización documental de brechas	Indicadores agregados	ANC por personal calificado, ANC4+, ANC8+, momento del primer contacto, área, departamento, etnia, educación	Usada
OMS, recomendaciones ANC [5]	Marco normativo	Guía clínica internacional	Ocho contactos prenatales, primer contacto en primer trimestre, calidad del cuidado	Usada
PAHO/PLISA [2]	Contexto sanitario nacional	Indicadores agregados	Mortalidad materna, indicadores de salud materna	Complementaria

HCP/SIP-CLAP [15]	Fuente para validación futura	Registro clínico institucional	Variables obstétricas, antecedentes, controles, riesgo materno, seguimiento	Requiere autorización institucional
HCP-SESAL	Fuente para piloto futuro	Registro clínico institucional	Datos prenatales rutinarios, establecimiento, seguimiento, referencia	Requiere aprobación ética y técnica
TRIPOD+AI / PROBAST+AI [22], [23]	Marco de validación futura	Guías metodológicas	Reporte, riesgo de sesgo, aplicabilidad, análisis del modelo	Usada como referencia metodológica

C. *Variables de análisis*

Las variables propuestas para el modelo predictivo se organizan en cinco dominios, con referencia explícita a la Historia Clínica Perinatal del CLAP/OPS [15], que es el instrumento estandarizado en uso dentro del sistema de la Secretaría de Salud de Honduras (SESAL). El dominio obstétrico incluye número de visitas ANC previas, semanas de gestación al primer contacto, paridad, antecedente de pérdida gestacional y condición de riesgo obstétrico registrada. El dominio sociodemográfico comprende edad materna, nivel educativo, estado civil, condición de aseguramiento y pertenencia étnica autodeclarada. El dominio geográfico abarca departamento, municipio, zona rural o urbana y distancia estimada a la unidad de salud más cercana. El dominio del sistema de salud incluye establecimiento de adscripción, nivel de resolución, disponibilidad de ultrasonido y tiempo de espera reportado.

Se proponen tres desenlaces binarios complementarios. El primero será el inicio tardío de la atención prenatal, definido como primer contacto después de las doce semanas de gestación, conforme al modelo de atención prenatal de la OMS [5]. El segundo será el incumplimiento de ANC4+, definido como menos de cuatro contactos prenatales durante el embarazo, indicador reportado por ENDESA/MICS [3]. El tercero será el incumplimiento de ANC8+, definido como menos de ocho contactos prenatales, de acuerdo con el estándar ampliado recomendado por la OMS y operacionalizado en ENDESA/MICS 2019 [5]-[3].

D. *Pipeline de modelado XAI*

El *pipeline* propuesto comprende cinco etapas (ver Fig. 1). La Etapa 1 comprende adquisición y gobernanza de datos mediante integración de la Historia Clínica Perinatal digital (HCP) con datos geográficos y del sistema SESAL, con uso propuesto de estándares FHIR R4 para interoperabilidad. La Etapa 2 incluye preprocesamiento con imputación de datos faltantes mediante KNN o regresión múltiple según el patrón de "missingness", codificación de variables categóricas, normalización *min-max* y detección de sesgos algorítmicos por grupo étnico y zona geográfica. La Etapa 3 contempla el desarrollo comparativo de modelos inherentemente interpretables y modelos de mayor capacidad predictiva. Se evaluarán regresión logística penalizada y árboles de decisión CART como modelos de referencia interpretables, junto con Random Forest y XGBoost como modelos no lineales de comparación [16], [17]. La selección del modelo no dependerá únicamente del AUC-ROC, sino también de sensibilidad, especificidad, F1-score, calibración, estabilidad explicativa, desempeño por subgrupos y factibilidad operativa. La Etapa 4 incorporará explicabilidad global y local mediante SHAP y TreeSHAP para estimar la contribución de las variables al riesgo de discontinuidad prenatal [18], [19]. LIME se utilizará como método complementario de explicación local y análisis de consistencia explicativa en casos seleccionados [20]. Las reglas derivadas de modelos interpretables, especialmente árboles de decisión se usarán como insumo para comunicación clínica y capacitación del personal del primer nivel de atención. La Etapa 5 contempla despliegue y monitoreo con interfaz del *dashboard* para personal de salud, sistema de alertas integrado al expediente clínico y protocolo de auditoría de equidad periódica.

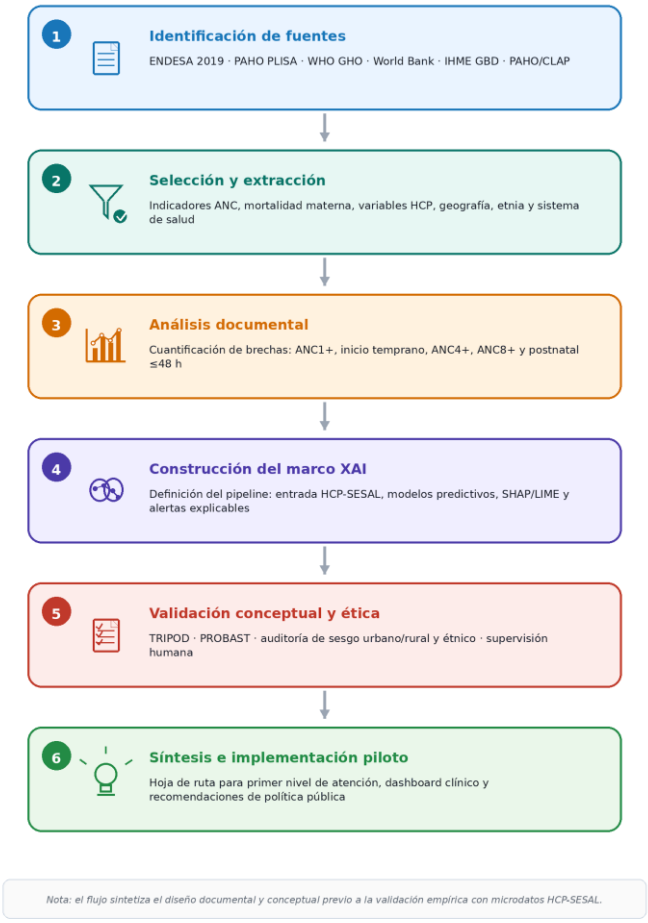


Fig. 1 Flujo metodológico del estudio. El esquema resume las cinco etapas de la propuesta: identificación de fuentes secundarias, extracción de indicadores, caracterización de brechas del continuo prenatal, diseño del pipeline XAI y definición del protocolo de validación futura.

E. *Fase de simulación del pipeline*

Para ilustrar la viabilidad operativa del pipeline en ausencia

de microdatos, se generó una cohorte sintética de 20,000 gestaciones cuyos parámetros marginales se calibraron con los indicadores publicados de ENDESA/MICS 2019: distribución urbano-rural, cobertura ANC8+ por área, por departamento (Distrito Central, Olancho y Gracias a Dios) y en población misquita, y proporción de primer contacto antes de los tres meses [3]. Los predictores reprodujeron los dominios de la Sección II-C: zona, departamento, pertenencia étnica, edad, escolaridad, paridad, aseguramiento, distancia a la unidad de salud e inicio tardío. Las relaciones individuales entre predictores y desenlace, que la encuesta no publica, se especificaron mediante un modelo logístico con efectos de magnitud moderada, un término no lineal para distancias superiores a 10 km en zona rural y una interacción entre inicio tardío y edad menor de 20 años; los interceptos por área y territorio se ajustaron de forma iterativa hasta reproducir las coberturas de referencia. Sobre el desenlace de incumplimiento de ANC8+ se ejecutaron las etapas 2 a 4 del pipeline: partición estratificada 70/30; entrenamiento de regresión logística con penalización L2 y validación cruzada, CART de profundidad 4, Random Forest y XGBoost; evaluación de discriminación (AUC con IC95% por bootstrap de 500 réplicas), puntuación de Brier, intercepto y pendiente de calibración, y sensibilidad, especificidad y F1 con umbral de 0.5; análisis por subgrupos; y explicabilidad global y local con TreeSHAP [16]–[19]. Se empleó Python 3.12 con scikit-learn 1.8, XGBoost 3.4 y SHAP 0.52. Los resultados de esta fase ilustran el funcionamiento del flujo analítico y no constituyen estimaciones del riesgo real en la población hondureña.

F. Consideraciones éticas

El presente trabajo se basa exclusivamente en análisis documental de información secundaria, agregada y públicamente disponible, por lo que no involucra recolección directa de datos individuales, identificables o sensibles de seres humanos. En consecuencia, esta fase metodológica no requiere consentimiento informado individual ni aprobación de comité de ética. No obstante, cualquier implementación futura del *pipeline* con microdatos del proveedor de servicio de salud (HCP-SESAL), Sistema Informático Perinatal / Centro Latinoamericano de Perinatología, Salud de la Mujer y Reproductiva (SIP/CLAP) o registros clínicos institucionales deberá contar con aprobación de un comité de ética en investigación, autorización técnica de SESAL, desidentificación previa de los datos, control de acceso, trazabilidad de uso y mecanismos explícitos de supervisión humana. La validación futura deberá incorporar evaluación de riesgo de sesgo y aplicabilidad conforme a PROBAST+AI, así como reporte transparente conforme a TRIPOD+AI [22], [23]. El modelo no deberá sustituir la decisión clínica, sino operar como herramienta de apoyo para priorización, seguimiento comunitario y gestión territorial del riesgo.

III. Desarrollo de la propuesta metodológica y resultados esperados

Dado que el presente trabajo no utiliza microdatos individuales ni entrena modelos predictivos sobre registros clínicos reales, los hallazgos se presentan como productos metodológicos derivados del análisis documental y del diseño del *pipeline* XAI. Por tanto, los resultados corresponden a: i) caracterización de brechas documentadas en fuentes secundarias; ii) definición de variables candidatas; iii) arquitectura analítica propuesta; iv) simulación del pipeline con una cohorte sintética; v) protocolo de validación piloto; y vi) indicadores de impacto sobre la continuidad prenatal.

A. Brechas documentadas en el continuo prenatal de Honduras

El análisis de ENDESA/MICS 2019 permite identificar una pérdida progresiva de cobertura a lo largo del continuo prenatal. La atención inicial por personal de salud calificado fue alta a nivel nacional, con 96.1%; sin embargo, 3.7% de las mujeres no tuvo ninguna visita prenatal y 7.5% tuvo únicamente entre una y tres visitas. La cobertura ANC4+ alcanzó 88.3%, mientras que la cobertura ANC8+ descendió a 39.0% [3]. Esta diferencia indica que el principal problema no es la entrada inicial al sistema, sino la continuidad suficiente del seguimiento prenatal conforme al estándar ampliado de la OMS.

Las brechas se acentúan territorialmente, ver Fig. 2. ENDESA/MICS 2019 reportó ANC8+ de 45.8% en zona urbana y 33.8% en zona rural. A nivel departamental, Gracias a Dios presentó 64.3% de ANC4+ y 20.0% de ANC8+, mientras que Olancho presentó 74.1% de ANC4+ y 20.5% de ANC8+. Estos valores contrastan con el Distrito Central, donde ANC8+ alcanzó 62.6% [3]. También se observan diferencias por pertenencia étnica del jefe del hogar: en población Misquita, ANC4+ fue 63.2% y ANC8+ 16.4%, lo que sugiere una brecha crítica de continuidad prenatal en territorios y grupos históricamente vulnerabilizados [3]. Ver Tabla II con detalles pormenorizados.

B. Marco XAI propuesto

El componente central de la propuesta es la integración de predicción e interpretabilidad en un flujo operativo aplicable al primer nivel de atención. En una implementación futura, los valores SHAP permitirían identificar la contribución media de variables clínicas, sociodemográficas, geográficas y del sistema de salud al riesgo de discontinuidad prenatal [18], [20]. Sin embargo, dado que el presente estudio no entrena modelos con microdatos individuales, no se reportan pesos SHAP reales ni jerarquías empíricas de predictores. los valores SHAP de la Sección III-D provienen de la cohorte sintética y solo verifican el funcionamiento del componente explicativo. En esta fase, la utilidad de SHAP y LIME se plantea como parte del diseño metodológico y no como resultado analítico ya validado [18]–[21].

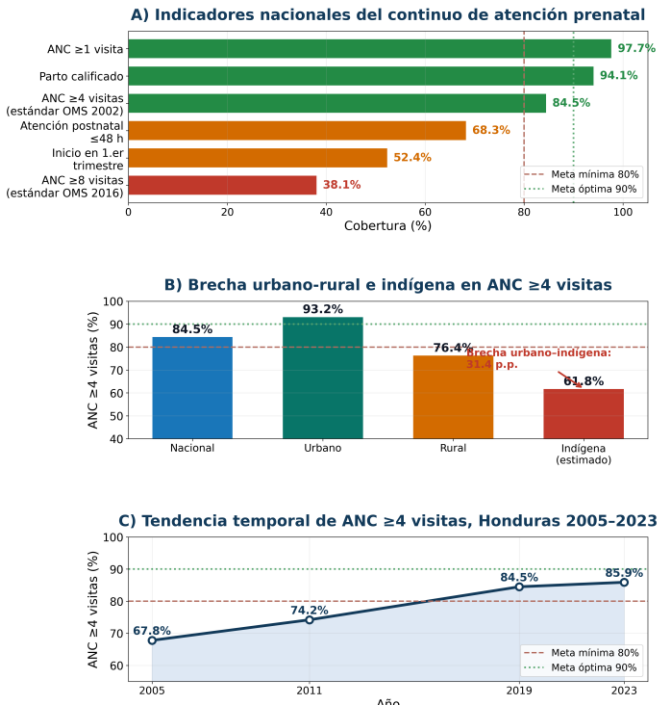


Fig. 2 Brechas documentadas en el continuo de atención prenatal en Honduras. Panel A: pérdida progresiva entre atención prenatal inicial, ANC4+ y ANC8+. Panel B: comparación urbano-rural de ANC4+ y ANC8+. Panel C: variación departamental de ANC8+, destacando los valores más bajos en Gracias a Dios y Olancho frente al Distrito Central. ANC = atención prenatal; p.p. = puntos porcentuales. Fuente: ENDESA/MICS 2019 [3].

En el nivel local, la explicación para una gestante individual tomará la forma de una lista de factores modificables ordenados por magnitud de contribución al riesgo, facilitando la consejería personalizada por parte del personal de salud de primer nivel [9]. La Fig. 3 muestra la arquitectura completa del pipeline XAI propuesto.

TABLA II
INDICADORES DEL CONTINUO DE ATENCIÓN PRENATAL (HONDURAS)

Indicador	Nacional (%)	Urbana (%)	Rural (%)	Brecha
Atención prenatal ≥1 vez por personal calificado	96.1	97.6	95.0	Gracias a Dios: 83.6; Olancho: 89.1
ANC ≥4 visitas	88.3	90.4	86.7	Gracias a Dios: 64.3; Olancho: 74.1
ANC ≥8	39.0	45.8	33.8	Gracias a Dios: 20.0; Olancho: 20.5
Primera visita antes de 4 meses	80.6	82.8	79.0	Gracias a Dios: 60.7; Olancho: 68.8
Primera visita antes de 3 meses	65.9	70.8	62.2	Gracias a Dios: 40.3; Olancho: 53.5
ANC ≥8 visitas en población Misquita	16.4	--	--	Brecha por pertenencia étnica del jefe del hogar

Nota: ANC = atención prenatal; el guion indica que el informe no presenta ese valor con la misma estructura urbano/rural para el subgrupo indicado. Los valores corresponden a mujeres de 15 a 49 años con un hijo nacido vivo en los dos años previos a la encuesta [3].

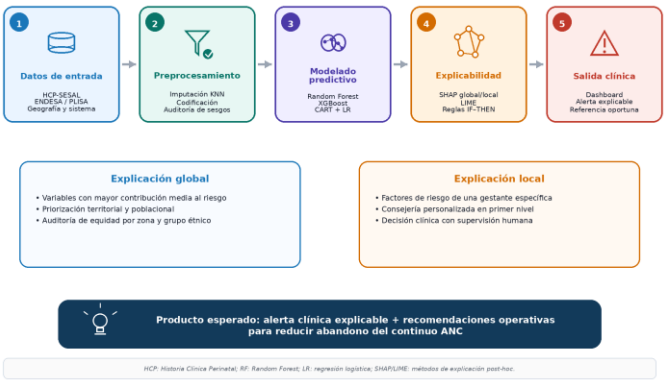


Fig. 3 Arquitectura XAI propuesta para la detección de brechas en el continuo de atención prenatal en Honduras. El pipeline integra variables clínicas, sociodemográficas, geográficas y del sistema de salud procedentes de la Historia Clínica Perinatal y de registros institucionales. La arquitectura contempla preprocesamiento, modelado predictivo, explicabilidad global y local mediante SHAP/LIME, generación de alertas clínicas y auditoría periódica de equidad.

C. Alineación con el sistema de información perinatal de Honduras

La Historia Clínica Perinatal y el Sistema Informático Perinatal del CLAP/OPS constituyen instrumentos regionales ampliamente utilizados para estandarizar el registro de información materna y perinatal [15]. La propuesta se alinea con esta estructura porque las variables candidatas del modelo como ser edad materna, antecedentes obstétricos, edad gestacional, controles prenatales, riesgo obstétrico y datos del establecimiento; son compatibles con registros clínicos prenatales rutinarios. Esta alineación reduce la necesidad de instrumentos adicionales y aumenta la factibilidad de una validación piloto en servicios del primer nivel de atención.

D. Simulación del pipeline con una cohorte sintética

La cohorte sintética reprodujo los indicadores de referencia con diferencias absolutas de entre 0.1 y 2.6 puntos porcentuales (Tabla III). La cobertura ANC8+ simulada fue de 38.7% a nivel nacional (referencia: 39.0%), 45.0% en zona urbana (45.8%), 33.9% en zona rural (33.8%), 20.4% en Olancho (20.5%) y 17.6% en población misquita (16.4%), lo que confirma que la simulación conserva el gradiente territorial y étnico documentado por la encuesta.

TABLA III
FIDELIDAD DE LA COHORTE SINTÉTICA FRENTE A ENDESA/MICS 2019

Indicador	ENDESA/MICS 2019 (%)	Cohorte sintética (%)
ANC8+ nacional	39.0	38.7
ANC8+ urbana	45.8	45.0
ANC8+ rural	33.8	33.9
ANC8+ Distrito Central	62.6	61.5
ANC8+ Olancho	20.5	20.4
ANC8+ Gracias a Dios	20.0	20.7
ANC8+ población misquita	16.4	17.6
Primer contacto <3 meses, nacional	65.9	66.6
Primer contacto <3 meses, Gracias a Dios	40.3	41.5

En el conjunto de prueba (n=6,000), la regresión logística penalizada, Random Forest y XGBoost alcanzaron una discriminación equivalente (AUC 0.727; IC95%: 0.715–0.740), superior a la del árbol CART (0.705; IC95%: 0.693–0.717) (Tabla IV). Los interceptos de calibración fueron próximos a cero y las pendientes se situaron entre 0.99 y 1.05, salvo en Random Forest, cuya pendiente de 1.22 indica predicciones comprimidas hacia la media (Fig. 4A). Con el umbral de 0.5, el árbol CART maximizó la sensibilidad (0.94) a costa de una especificidad de 0.20, insuficiente para uso operativo. Que los modelos complejos no superaran a la regresión logística es coherente con la estructura predominantemente aditiva de la simulación y muestra por qué la selección no debe basarse en la complejidad del algoritmo, sino en una ganancia de desempeño demostrada [9].

TABLA IV
DESEMPEÑO DE LOS MODELOS EN LA SIMULACIÓN (CONJUNTO DE PRUEBA)

Modelo	AUC (IC95%)	Brier	Pendiente de calibración	Sens. / Esp.
Regresión logística L2	0.727 (0.715–0.740)	0.202	1.05	0.82 / 0.46
Árbol CART	0.705 (0.693–0.717)	0.206	1.03	0.94 / 0.20
Random Forest	0.727 (0.715–0.740)	0.203	1.22	0.84 / 0.42
XGBoost	0.727 (0.715–0.739)	0.202	0.99	0.81 / 0.48

Nota: IC95% por bootstrap (500 réplicas); sensibilidad y especificidad con umbral de 0.5. Datos sintéticos.

El análisis por subgrupos no mostró deterioro en zona rural (AUC de XGBoost 0.725) frente a la urbana (0.709) ni en adolescentes (0.735), y la razón observada/esperado se mantuvo entre 0.98 y 1.05. Sin embargo, en Gracias a Dios (n=77) y en población misquita (n=55) los intervalos de confianza fueron amplios (0.56–0.93 en esta última), de modo que la exactitud del modelo en los grupos con mayor brecha no pudo establecerse con precisión. Este resultado anticipa un requisito del piloto: sin sobremuestreo deliberado de territorios indígenas y rurales dispersos, la auditoría de equidad carece de potencia estadística.

La explicabilidad global con TreeSHAP situó el inicio tardío, la distancia a la unidad de salud y la zona rural como principales contribuyentes al riesgo (Fig. 4B), orden concordante con la estructura con que se generaron los datos, lo que verifica que el componente explicativo recupera las relaciones especificadas. En el nivel local, para una gestante adolescente de zona rural de Olancho, con primer contacto tardío y residencia a 29.6 km del establecimiento, el modelo estimó un riesgo de 0.98 y la explicación priorizó distancia, inicio tardío, territorio y edad; en la práctica, esa información orientaría la visita domiciliaria, el apoyo para transporte y la programación anticipada de citas.

E. Protocolo de validación piloto e indicadores de impacto

La validación piloto se estructuró en cuatro fases secuenciales. La primera es una validación retrospectiva con datos anonimizados de la HCP-SESAL procedentes de establecimientos del primer nivel en al menos un departamento de alta brecha y otro de referencia, con partición temporal (entrenamiento con embarazos de años previos y prueba con el año más reciente) para aproximar el uso prospectivo. En ella se repetirá la comparación entre modelos interpretables y complejos ejecutada en la simulación; un modelo complejo solo se adoptará si mejora la discriminación y el beneficio neto del análisis de curva de decisión sin deteriorar la calibración ni la equidad [9], [22], [23], [24]. La calibración se evaluará con intercepto, pendiente y curvas suavizadas, con recalibración local cuando proceda, y la equidad mediante AUC, sensibilidad y razón observado/esperado por pertenencia étnica, zona, departamento y edad, con sobremuestreo de los grupos pequeños.

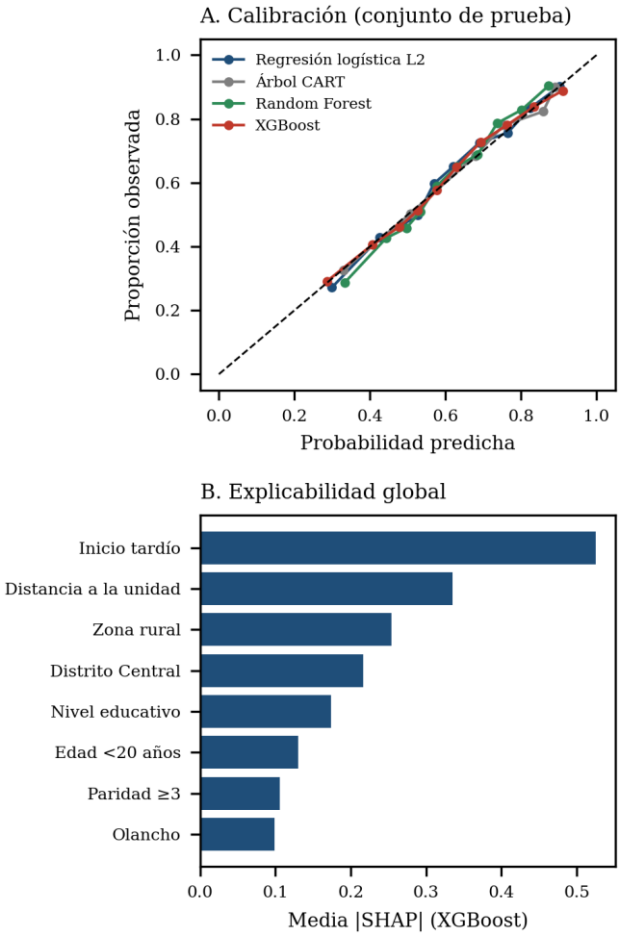


Fig. 4 Simulación del pipeline XAI con una cohorte sintética calibrada con ENDESA/MICS 2019 (n=20,000; conjunto de prueba n=6,000). Panel A: curvas de calibración por deciles de probabilidad predicha para los cuatro modelos. Panel B: importancia global de los ocho predictores principales según la media del valor absoluto de SHAP en XGBoost. Resultados ilustrativos, no estimaciones empíricas.

La segunda fase incorpora a los usuarios finales. Mediante un panel Delphi con personal médico, de enfermería y promotores de salud se definirán el umbral de alerta, el formato de la explicación local y las acciones asociadas a cada nivel de riesgo; en paralelo, grupos focales con gestantes y puérperas de zonas urbanas, rurales e indígenas valorarán la aceptabilidad de la herramienta, el lenguaje de la consejería y el riesgo percibido de estigmatización. La tercera fase es un despliegue silencioso, en el que el modelo genera alertas que no se muestran al personal, para contrastar su desempeño con los desenlaces observados antes de cualquier uso clínico. La cuarta es un ensayo por conglomerados con diseño escalonado (stepped-wedge) entre establecimientos, que permite estimar el impacto de las alertas sobre la continuidad prenatal mientras todos los establecimientos acceden progresivamente a la intervención. Los indicadores y criterios propuestos para cada nivel de evaluación se resumen en la Tabla V.

TABLA V
INDICADORES PARA EVALUAR EL DESEMPEÑO Y EL IMPACTO DEL SISTEMA

Nivel	Indicador	Criterio propuesto
Modelo	AUC y beneficio neto (curva de decisión)	AUC ≥0.70; beneficio neto superior a intervenir a todas o a ninguna
Modelo	Intercepto y pendiente de calibración, global y por subgrupo	Pendiente 0.8–1.2; razón O/E 0.9–1.1
Equidad	Diferencia de sensibilidad entre grupos étnicos y zonas	≤10 puntos porcentuales
Proceso	Alertas con acción documentada en ≤7 días	≥80%
Proceso	Aceptabilidad para personal y gestantes	≥75% de valoraciones favorables
Resultado	Primer contacto ≤12 semanas	Aumento frente a conglomerados control
Resultado	Cobertura ANC8+	Aumento frente a conglomerados control
Resultado	Gestantes reincorporadas tras inasistencia	Proporción recuperada en ≤30 días
Resultado	Brecha rural-urbana e indígena en ANC8+	Reducción respecto de la línea base

IV. DISCUSIÓN

Los indicadores documentados evidencian una doble brecha en el continuo prenatal hondureño: una brecha de oportunidad, relacionada con el inicio tardío de la atención prenatal, y una brecha de continuidad, reflejada en la caída entre ANC4+ y ANC8+. Aunque la atención inicial por personal calificado es alta, la proporción de mujeres que alcanza ocho o más contactos prenatales desciende a 39.0%, con valores sustancialmente menores en zonas rurales, Gracias a Dios, Olancho y población Misquita [3]. Este patrón sugiere que las intervenciones no deben concentrarse únicamente en captar gestantes, sino en sostener el seguimiento longitudinal durante todo el embarazo.

Desde el punto de vista metodológico, la propuesta debe evitar depender exclusivamente de modelos de caja negra explicados de forma post hoc. En decisiones clínicas o de salud pública con impacto sobre poblaciones vulnerables, es recomendable comparar modelos inherentemente interpretables con modelos de mayor desempeño predictivo [9]. Por ello, el pipeline propuesto incorpora regresión logística penalizada y árboles de decisión como línea base interpretable, y Random Forest y XGBoost como modelos comparativos de mayor capacidad no lineal [16], [17]. La decisión final de implementación deberá basarse en una evaluación balanceada de desempeño, calibración, interpretabilidad, estabilidad explicativa y equidad entre subgrupos [22], [23].

El marco XAI propuesto responde precisamente a esta doble brecha. A diferencia de los sistemas de alerta basados en reglas fijas (que ignoran la variabilidad individual), los modelos de aprendizaje automático pueden capturar interacciones no lineales entre los factores de riesgo [7]. Sin embargo, sin explicabilidad, su utilidad clínica sería limitada: el personal de salud de primer nivel necesita entender por qué un modelo predice riesgo para una gestante específica, no solo saber que lo predice [9]. Este es el aporte central de la capa SHAP/LIME: traducir la predicción en una lista de factores modificables que el proveedor de salud pueda abordar en la consulta.

La propuesta se inscribe en una tendencia global de aplicación de XAI en salud materna. Holzinger et al. [12] argumentan que la explicabilidad es una condición necesaria, no suficiente, para la confianza en sistemas de IA clínica, y que la causalidad explícita debe ser el objetivo final más allá de la correlación. Arrieta et al. [11] proponen una taxonomía de métodos XAI que sitúa a SHAP en la categoría post-hoc agnóstica al modelo, con ventajas en consistencia matemática (cumple axiomas de Shapley) y desventajas en costo computacional con datasets muy grandes. LIME, por su parte, ofrece mayor velocidad a expensas de menor estabilidad local [20]. La combinación de ambos métodos en el pipeline propuesto responde a esta complementariedad.

Una fortaleza específica del marco es su alineación con la Historia Clínica Perinatal del CLAP/OPS [15], ya adoptada en Honduras. Esto evita el problema de fragmentación de sistemas, frecuente en implementaciones de IA en salud de bajos y medianos ingresos [25]. La posibilidad de desplegar el modelo como módulo adicional del HCP digital, con alertas integradas al flujo de trabajo del personal de salud, reduce las barreras de adopción identificadas en contextos similares de Centroamérica.

No obstante, el marco enfrenta limitaciones que deben reconocerse. La calidad del HCP en zonas rurales e indígenas es variable y el subregistro puede introducir sesgos sistemáticos en el modelo precisamente en las poblaciones más vulnerables [5]. La validación del modelo requerirá una muestra mínima de varios miles de registros individuales; la negociación de acceso a los microdatos de SESAL o a los datos anonimizados del HCP será el cuello de botella operativo de la fase empírica. La simulación mostró, además, que incluso con 20,000 registros la población misquita queda

representada por pocas decenas de casos en el conjunto de prueba, lo que obliga a planificar el muestreo en función de la auditoría de equidad y no solo del tamaño total. La equidad algorítmica debe ser monitoreada de forma continua: los modelos entrenados con datos históricos de un sistema inequitativo pueden perpetuar o amplificar las desigualdades existentes [13], riesgo especialmente relevante cuando las poblaciones indígenas están subrepresentadas en los datos de entrenamiento.

Desde la perspectiva de política pública, el marco propuesto es consistente con el Plan Nacional de Salud 2021-2026 de SESAL [25] y de UNICEF [26] que prioriza la reducción de la mortalidad materna mediante el fortalecimiento del primer nivel de atención y la adopción de tecnologías digitales. La adopción de un sistema XAI en el primer nivel requerirá capacitación del personal, infraestructura de conectividad y un protocolo claro de referencia para gestantes con riesgo identificado de abandono del continuo.

Este estudio tiene limitaciones importantes. Primero, la propuesta no incluye entrenamiento ni validación empírica de modelos predictivos, debido a la falta de acceso a microdatos individuales del sistema HCP-SESAL o de ENDESA/MICS. La simulación de la Sección III-D demuestra la ejecución técnica del pipeline, pero sus métricas dependen de supuestos sobre la estructura individual del riesgo que la encuesta no pública y no deben interpretarse como el desempeño esperado en registros reales. Segundo, los indicadores utilizados corresponden a fuentes agregadas y, por tanto, no permiten estimar asociaciones individuales ni inferir causalidad. Tercero, algunas variables críticas para explicar la discontinuidad prenatal, como distancia efectiva al establecimiento, costos indirectos, calidad percibida de la atención y disponibilidad real de transporte, podrían no estar registradas de forma sistemática en los sistemas actuales. Cuarto, cualquier implementación futura deberá considerar sesgo de selección, subregistro diferencial en zonas rurales e indígenas, calibración por subgrupos y supervisión humana obligatoria antes de utilizar alertas predictivas en la práctica clínica.

Por tanto, el marco debe presentarse como una propuesta metodológica preempírica y no como un sistema predictivo validado. Esta distinción es esencial para evitar sobreinterpretación de los resultados y para alinear el manuscrito con los estándares actuales de reporte y evaluación de modelos clínicos basados en IA [22], [23].

V. Conclusiones

El continuo de atención prenatal en Honduras muestra pérdidas progresivas entre la atención inicial por personal calificado, el inicio oportuno y el cumplimiento del estándar ampliado de ocho contactos. ENDESA/MICS 2019 evidencia que, aunque la cobertura inicial es alta, solo 39.0% de las mujeres alcanza ANC8+, con brechas marcadas en zonas rurales, departamentos periféricos y población Misquita.

El marco XAI propuesto constituye una arquitectura metodológica para transformar registros prenatales rutinarios en alertas explicables de riesgo de discontinuidad, integrando modelos predictivos, explicabilidad global y local, auditoría de equidad y supervisión humana. Su contribución principal no consiste en presentar un modelo ya entrenado, sino en ofrecer una ruta técnica para una futura validación con microdatos anonimizados del sistema prenatal hondureño. La simulación con una cohorte calibrada con ENDESA/MICS 2019 mostró que el flujo es ejecutable de extremo a extremo, que un modelo interpretable puede igualar el desempeño de los complejos y que la auditoría de equidad exige sobremuestrear los grupos con mayor brecha.

La siguiente fase deberá incluir acceso autorizado a microdatos, definición prospectiva de desenlaces, validación interna y externa, evaluación de calibración, análisis de equidad por subgrupos, prueba piloto en establecimientos seleccionados co-diseño con personal sanitario y gestantes, despliegue silencioso, ensayo escalonado con indicadores de impacto sobre la continuidad prenatal y reporte conforme a TRIPOD+AI y PROBAST+AI.

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, “Maternal mortality”, *WHO Fact Sheet*, Apr. 7, 2025.

[2] Pan American Health Organization, “Core Indicators: Health Situation in the Americas”, *PAHO/EIH Open Data*, 2025. [Online]. Available: https://opendata.paho.org/en/core-indicators. Accessed: May 4, 2026.

[3] 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*. Tegucigalpa, Honduras: INE, 2020. [Online]. Available: https://ine.gob.hn/2020/06/17/informe-endesa-mics-2019/. Accessed: May 4, 2026.

[4] World Health Organization, UNICEF, UNFPA, World Bank Group, and UNDESA/Population Division, *Trends in Maternal Mortality 2000 to 2020: Estimates by WHO, UNICEF, UNFPA, World Bank Group and UNDESA/Population Division*. Geneva, Switzerland: World Health Organization, 2023. [Online]. Available: https://www.who.int/publications/i/item/9789240068759. Accessed: May 4, 2026.

[5] World Health Organization, *WHO Recommendations on Antenatal Care for a Positive Pregnancy Experience*. Geneva, Switzerland: World Health Organization, 2016. [Online]. Available: https://www.who.int/publications/i/item/9789241549912. Accessed: May 4, 2026

[6] Ö. Tunçalp et al., “Quality of care for pregnant women and newborns — The WHO vision”, *BJOG*, vol. 122, no. 8, pp. 1045–1049, Jul. 2015, doi: 10.1111/1471-0528.13451.

[7] E. J. Topol, “High-performance medicine: The convergence of human and artificial intelligence”, *Nat. Med.*, vol. 25, no. 1, pp. 44–56, Jan. 2019, doi: 10.1038/s41591-018-0300-7
[8] Z. Obermeyer and E. J. Emanuel, “Predicting the future — Big data, machine learning, and clinical medicine”, *N. Engl. J. Med.*, vol. 375, no. 13, pp. 1216–1219, Sep. 2016, doi: 10.1056/NEJMp1606181
[9] C. Rudin, “Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead”, *Nat. Mach. Intell.*, vol. 1, no. 5, pp. 206–215, May 2019, doi: 10.1038/s42256-019-0048-x.
[10] Z. C. Lipton, “The mythos of model interpretability”, *Queue*, vol. 16, no. 3, pp. 31–57, Jun. 2018, doi: 10.1145/3236386.3241340.
[11] A. Barredo Arrieta et al., “Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI”, *Inf. Fusion*, vol. 58, pp. 82–115, Jun. 2020, doi: 10.1016/j.inffus.2019.12.012
[12] A. Holzinger, G. Langs, H. Denk, K. Zatloukal, and H. Müller, “Causability and explainability of artificial intelligence in medicine”, *WIREs Data Min. Knowl. Discov.*, vol. 9, no. 4, Art. no. e1312, Jul./Aug. 2019, doi: 10.1002/widm.1312.
[13] World Health Organization, *Ethics and Governance of Artificial Intelligence for Health: WHO Guidance*. Geneva, Switzerland: WHO, 2021.
[14] D. S. Char, N. H. Shah, and D. Magnus, “Implementing machine learning in health care — Addressing ethical challenges”, *N. Engl. J. Med.*, vol. 378, no. 11, pp. 981–983, 2018, doi: 10.1056/NEJMp1714229.
[15] Pan American Health Organization, “Sistema Informático Perinatal (SIP)”, Centro Latinoamericano de Perinatología, Salud de la Mujer y Reproductiva, PAHO/WHO, 2024.
[16] L. Breiman, “Random forests”, *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.
[17] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system”, in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, San Francisco, CA, USA, 2016, pp. 785–794, doi: 10.1145/2939672.2939785.
[18] S. M. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions”, in *Adv. Neural Inf. Process. Syst.* Vol. 30, 2017, pp. 4765–4774.
[19] S. M. Lundberg et al., “From local explanations to global understanding with explainable AI for trees”, *Nat. Mach. Intell.*, vol. 2, no. 1, pp. 56–67, Jan. 2020, doi: 10.1038/s42256-019-0138-9
[20] M. T. Ribeiro, S. Singh, and C. Guestrin, “‘Why should I trust you?’: Explaining the predictions of any classifier”, in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, San Francisco, CA, USA, 2016, pp. 1135–1144, doi: 10.1145/2939672.2939778.
[21] HL7 International, “FHIR Release 4”, Health Level Seven International, 2019.
[22] 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.
[23] K. G. M. Moons et al., “PROBAST+AI: An updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods,” *BMJ*, vol. 388, Art. no. e082505, Mar. 2025, doi: 10.1136/bmj-2024-082505.
[24] A. J. Vickers and E. B. Elkin, “Decision curve analysis: A novel method for evaluating prediction models”, *Med. Decis. Making,* vol. 26, no. 6, pp. 565–574, Nov. 2006, doi: 10.1177/0272989X06295361.
[25] Secretaría de Estado en el Despacho de Salud, *Plan Estratégico Institucional 2023–2026*. Tegucigalpa, Honduras: Secretaría de Salud, 2023. [Online]. Available: https://p4h.world/app/uploads/2025/03/Plan-estrategico-institucional_2023-%E2%80%93-2026-de-salud_-Honduras.x73677.pdf. Accessed: May 4, 2026.
[26] UNICEF, “Antenatal care”, *UNICEF Data*, Nov. 2024. [Online]. Available: https://data.unicef.org/topic/maternal-health/antenatal-care/. Accessed: May 4, 2026.