

Predictive Modeling of Child Stunting in Honduras Using Explainable Machine Learning

Yolly Molina¹; Jorge Valle-Reconco²; Patricia Soriano³; Arnoldo Zelaya⁴

¹Centro de Diagnóstico de Imágenes Biomédicas, Investigación y Rehabilitación, Universidad Nacional Autónoma de Honduras, Honduras, yolly.molina@unah.edu.hn

^{2,4}Facultad de Ciencias Médicas, Universidad Nacional Autónoma de Honduras, Honduras, jorge_valle@unah.edu.hn, arnoldo.zelaya@unah.edu.hn

³Dirección de Medicina Forense, Ministerio Público, Honduras, paty.soriano@yahoo.com

Abstract— Chronic child malnutrition is a persistent public health problem in Honduras, with irreversible consequences on cognitive and physical development. The aim of this study was to develop and validate predictive models of chronic malnutrition (stunting) using explainable machine learning, to identify the most influential determinants and estimate their individual contribution to risk. Data from 8,713 children aged 0 to 59 months from ENDESA/MICS 2019 were analyzed. A wealth index was constructed using principal component analysis on 21 household assets and three algorithms (logistic regression, Random Forest and Gradient Boosting) were trained with stratified cross-validation. Interpretability was evaluated with SHAP (SHapley Additive exPlanations) values. The prevalence of stunting was 18.9%. Random Forest presented the best performance (AUC-ROC=0.691; AUC-PR=0.362). SHAP analyses identified wealth index as the primary predictor (mean SHAP=0.489), followed by child age (0.323) and maternal education (0.251). A marked socioeconomic gradient was observed (Q1: 39.2% vs Q5: 7.1%), with amplification of the effect in rural areas. In conclusion, explainable machine learning allows the identification and quantification of key determinants of chronic child malnutrition, supporting interventions focused on economic transfers, improvements in water and sanitation, and women's education, with territorial prioritization in the western corridor of the country.

Keywords— stunting, machine learning, SHAP, ENDESA/MICS, social determinants.

Modelado Predictivo de Desnutrición Crónica Infantil en Honduras Mediante Aprendizaje Automático Explicable

Yolly Molina¹; Jorge Valle-Reconco²; Patricia Soriano³; Arnoldo Zelaya⁴

¹Centro de Diagnóstico de Imágenes Biomédicas, Investigación y Rehabilitación, Universidad Nacional Autónoma de Honduras, Honduras, yolly.molina@unah.edu.hn

^{2,4}Facultad de Ciencias Médicas, Universidad Nacional Autónoma de Honduras, Honduras, jorge_valle@unah.edu.hn, arnoldo.zelaya@unah.edu.hn

³Dirección de Medicina Forense, Ministerio Público, Honduras, paty.soriano@yahoo.com

Resumen— La desnutrición crónica infantil constituye un problema persistente de salud pública en Honduras, con consecuencias irreversibles sobre el desarrollo cognitivo y físico. El objetivo de este estudio fue desarrollar y validar modelos predictivos de desnutrición crónica (stunting) mediante aprendizaje automático explicable, para identificar los determinantes más influyentes y estimar su contribución individual al riesgo. Se analizaron datos de 8,713 niños de 0 a 59 meses de la ENDESA/MICS 2019. Se construyó un índice de riqueza mediante análisis de componentes principales sobre 21 activos del hogar y se entrenaron tres algoritmos (regresión logística, Random Forest y Gradient Boosting) con validación cruzada estratificada. La interpretabilidad se evaluó con valores SHAP (SHapley Additive exPlanations). La prevalencia de stunting fue 18.9%. Random Forest presentó el mejor desempeño (AUC-ROC=0.691; AUC-PR=0.362). Los análisis SHAP identificaron el índice de riqueza como el predictor principal (SHAP medio=0.489), seguido por la edad del niño (0.323) y la educación materna (0.251). Se observó un gradiente socioeconómico marcado (Q1: 39.2% vs Q5: 7.1%), con amplificación del efecto en áreas rurales. En conclusión, el aprendizaje automático explicable permite identificar y cuantificar determinantes clave de la desnutrición crónica infantil, respaldando intervenciones focalizadas en transferencias económicas, mejoras en agua y saneamiento, y educación femenina, con priorización territorial en el corredor occidental del país.

Palabras clave— desnutrición crónica, aprendizaje automático, SHAP, ENDESA/MICS, determinantes sociales.

I. INTRODUCCIÓN

La desnutrición crónica infantil, manifestada como baja talla para la edad o stunting, constituye uno de los indicadores más sensibles de las condiciones de vida y desarrollo de una población. Según la Organización Mundial de la Salud (OMS), se define cuando el puntaje z de talla para edad (HAZ) se encuentra por debajo de menos dos desviaciones estándar respecto a los estándares de crecimiento infantil [1]. Este déficit antropométrico refleja una acumulación de episodios de desnutrición y enfermedades infecciosas durante los primeros 1,000 días de vida, período crítico conocido como la ventana de oportunidad para el desarrollo [2].

A nivel global, se estima que 148.1 millones de niños menores de cinco años padecen desnutrición crónica, lo que representa aproximadamente el 22.3% de esta población [3]. En América Latina y el Caribe, aunque la prevalencia promedio regional ha disminuido considerablemente en las

últimas décadas, persisten marcadas disparidades entre y dentro de los países. Guatemala presenta las cifras más elevadas del continente (46.5%), seguida de Ecuador (23.1%) y Honduras, donde las estimaciones recientes sitúan la prevalencia entre 17 y 23% dependiendo de la fuente y metodología empleada [4].

Las consecuencias del stunting trascienden el ámbito nutricional y afectan múltiples dimensiones del desarrollo humano. Los niños con desnutrición crónica presentan déficits cognitivos que persisten hasta la edad adulta, menor rendimiento escolar, reducida productividad laboral y mayor susceptibilidad a enfermedades crónicas no transmisibles como diabetes tipo 2 y enfermedades cardiovasculares [5]. El Banco Mundial estima que el stunting reduce el producto interno bruto de los países afectados entre 2 y 3%, perpetuando ciclos intergeneracionales de pobreza [6].

El marco conceptual de UNICEF sobre nutrición infantil establece que el stunting resulta de una compleja interacción entre causas inmediatas (ingesta alimentaria inadecuada y enfermedades), causas subyacentes (inseguridad alimentaria del hogar, prácticas de cuidado deficientes y acceso limitado a servicios de salud, agua y saneamiento) y causas básicas (pobreza, inequidad de género y bajo nivel educativo) [7]. Esta multicausalidad dificulta la identificación de los factores más relevantes para orientar intervenciones de política pública con recursos limitados.

En años recientes, las técnicas de aprendizaje automático (machine learning) han emergido como herramientas prometedoras para el análisis de determinantes de salud poblacional. A diferencia de los modelos estadísticos tradicionales, estos algoritmos pueden capturar relaciones no lineales y efectos de interacción complejos entre múltiples variables sin requerir especificaciones a priori [8]. Sin embargo, su naturaleza de caja negra ha limitado su adopción en el ámbito de la salud pública, donde la interpretabilidad de los resultados es crucial para informar decisiones políticas.

Los métodos de inteligencia artificial explicable (XAI), particularmente los valores SHAP (SHapley Additive exPlanations), ofrecen una solución a esta limitación al proporcionar explicaciones teóricamente fundamentadas sobre la contribución de cada variable a las predicciones del modelo [9]. Estudios recientes han aplicado estos enfoques para

identificar determinantes de desnutrición infantil en Bangladesh [10], Papua Nueva Guinea [11] y múltiples países africanos [12], demostrando su utilidad para revelar patrones ocultos en datos poblacionales complejos.

En Honduras, la evidencia sobre los determinantes del stunting proviene principalmente de análisis descriptivos y modelos de regresión logística tradicional aplicados a las Encuestas de Demografía y Salud. Estos estudios han identificado consistentemente la pobreza, la ruralidad, la baja educación materna y el limitado acceso a servicios básicos como factores de riesgo [13]. No obstante, hasta la fecha no se han aplicado técnicas de aprendizaje automático explicable que permitan cuantificar la importancia relativa de estos determinantes y caracterizar sus interacciones.

El presente estudio tiene como objetivo desarrollar y validar modelos predictivos de desnutrición crónica infantil utilizando datos de la Encuesta Nacional de Demografía y Salud de Honduras (ENDESA/MICS 2019), empleando técnicas de aprendizaje automático explicable para identificar los determinantes más influyentes, caracterizar sus relaciones funcionales con el outcome y evaluar posibles interacciones entre variables socioeconómicas y geográficas.

II. MATERIALES Y MÉTODOS

A. Fuente de datos y población de estudio

Se emplearon datos de la Encuesta Nacional de Demografía y Salud y de la Encuesta de Indicadores Múltiples por Conglomerados (ENDESA/MICS 2019) de Honduras, accesibles públicamente en el site del Instituto Nacional de Estadística [14]. Esta encuesta constituye el análisis poblacional más reciente y exhaustivo sobre salud materno-infantil en el país, con cobertura nacional y representatividad a nivel departamental. El trabajo de campo se llevó a cabo entre agosto de 2019 y enero de 2020, utilizando un diseño muestral bietápico estratificado con selección proporcional al tamaño de las unidades primarias de muestreo.

La población de estudio abarcó a todos los niños de 0 a 59 meses de edad con mediciones antropométricas válidas documentadas en el módulo de antropometría infantil. Se excluyeron los casos con valores implausibles de puntaje z de talla para edad, conforme a los criterios establecidos por la OMS ($HAZ < -6$ o $HAZ > 6$), así como aquellos con información ausente en las variables de interés que no pudo ser recuperada mediante la vinculación con otros módulos de la encuesta. La muestra final analizada consistió en 8,713 niños.

B. Variable dependiente

La variable dependiente fue la desnutrición crónica (stunting), definida como un puntaje z de talla por edad inferior a -2 desviaciones estándar en relación con los estándares de crecimiento infantil de la OMS 2006 [15]. Esta variable se codificó de forma binaria: presencia (1) o ausencia (0) de retraso en el crecimiento. El puntaje z fue determinado por la encuesta mediante el software ANTHRO de la OMS,

asegurando la comparabilidad internacional de las estimaciones.

C. Variable independiente

Se seleccionaron variables independientes conforme al marco conceptual de UNICEF para la nutrición infantil, organizadas en tres niveles: características del niño (edad en meses, sexo, morbilidad reciente definida como episodio de diarrea o infección respiratoria aguda en las dos semanas anteriores, práctica actual de lactancia materna y antecedentes de lactancia), características maternas (años de educación formal), y características del hogar y entorno (índice de riqueza, área de residencia rural/urbana, acceso a fuente mejorada de agua según la clasificación de la OMS, acceso a instalaciones de saneamiento mejorado, disponibilidad de electricidad y departamento de residencia).

D. Construcción del índice de riqueza

Debido a que la encuesta no incorpora mediciones directas de ingresos o gastos del hogar, se elaboró un índice de riqueza mediante un análisis de componentes principales (ACP) sobre un conjunto de 21 activos domésticos que abarca: posesión de televisor, radio, refrigerador, teléfono, computadora, vehículo, motocicleta, bicicleta, materiales predominantes de piso, paredes y techo de la vivienda, tipo de combustible para cocinar, tenencia de la vivienda, número de habitaciones para dormir y hacinamiento. Esta metodología, ampliamente empleada en las Encuestas de Demografía y Salud a nivel mundial [16], facilita la captura del nivel socioeconómico relativo de los hogares.

Las variables categóricas se convirtieron en indicadores binarios (dummies) antes del análisis. El primer componente mayor, que explicó el 28.3% de la varianza total, fue conservado como un índice de riqueza. Los puntajes obtenidos fueron clasificados en quintiles poblacionales, donde el quintil 1 representa el 20% más desfavorecido y el quintil 5 el 20% más favorecido. Se emplearon tanto la variable continua (puntaje de ACP) como la variable categórica (quintiles) en los modelos de aprendizaje automático.

E. Modelos de aprendizaje automático

Se entrenaron tres algoritmos de clasificación supervisada: la regresión logística con regularización L2 (Ridge) como un modelo de referencia interpretable, el Árbol Aleatorio como un método de ensamble fundamentado en árboles de decisión, y el Gradient Boosting como un algoritmo de mejora iterativa. La elección de dichos algoritmos se orientó hacia un equilibrio entre la interpretabilidad, la capacidad predictiva y la complementariedad en los supuestos subyacentes [17].

La muestra de datos fue subdividida aleatoriamente en dos grupos: entrenamiento (80%, $n=6,970$) y prueba (20%, $n=1,743$) a través de estratificación sobre la variable dependiente con el objetivo de mantener la prevalencia original. Los hiperparámetros de cada modelo fueron optimizados a través de la búsqueda en cuadrícula con

validación cruzada estratificada de cinco pliegues sobre el conjunto de formación. Para tratar el desequilibrio de clases, con una prevalencia del 18.9%, se implementó una ponderación de clases que se alineaba inversamente con su frecuencia.

F. Evaluación del desempeño predictivo

La evaluación del rendimiento predictivo en el conjunto de pruebas se realizó a través de diversas métricas complementarias. El área bajo la curva de capacidad discriminativa general (AUC-ROC) evaluó la capacidad discriminativa general del modelo. Debido al desequilibrio de clases, también se calculó el área bajo la curva precisión-recall (AUC-PR), que resulta más informativa cuando la clase positiva es minoritaria [18]. La puntuación proporcionada por Brier evaluó la calibración probabilística del modelo. Además, se documentaron la sensibilidad, la especificidad y el puntaje F1-utilizando el umbral de clasificación que optimiza el índice de Youden. Todo el procesamiento y análisis se realizó con software de código abierto (Python 3.11, scikit-learn 1.3, SHAP 0.42) [19]-[22].

G. Análisis de interpretabilidad con SHAP

Los valores SHAP (SHapley Additive exPlanations) fueron utilizados para la interpretación de los modelos, basándose en la teoría de juegos cooperativos [9]. Los valores SHAP descomponen la predicción de cada observación en aportaciones aditivas correspondientes a cada variable, cumpliendo con características deseadas como eficiencia local, consistencia y simetría. Para los modelos fundamentados en árboles, se empleó el algoritmo TreeSHAP, caracterizado por su eficiencia y precisión computacional [23].

Se produjeron tres categorías de visualizaciones SHAP: gráficos de resumen que evidencian la distribución de los valores SHAP para cada variable, gráficos de dependencia que evidencian la relación funcional entre el valor de una variable y su aportación a la predicción, y análisis de interacción que desvelan efectos sinérgicos o antagónicos entre pares de variables. Se llevó a cabo el análisis SHAP sobre el modelo Gradient Boosting debido a su calibración superior y su habilidad para capturar relaciones no lineales.

H. Consideraciones éticas

El presente estudio constituye un análisis secundario de datos de acceso público, anonimizados y sin posibilidad de identificación individual. La ENDESA/MICS 2019 fue aprobada por el Comité de Ética de ICF International y las autoridades sanitarias de Honduras. Los participantes proporcionaron consentimiento informado previo a su inclusión. El uso de los datos se realizó conforme a los términos de uso establecidos por el Instituto Nacional de Estadística de Honduras.

III. RESULTADOS

A. Características de la muestra

La población final comprendió 8,713 niños en el rango de edad de 0 a 59 meses, con información antropométrica válida. La prevalencia de desnutrición crónica se registró en un 18.9% (Intervalo de Confianza del 95%: 18.1-19.8%). La edad promedio registrada fue de 29.4 meses (Desviación estándar (DE): 17.2), con una distribución homogénea entre los géneros (50.9% masculino). El 64.5% de la población residía en zonas rurales. En relación con el acceso a servicios fundamentales, el 89.2% disponía de una fuente de agua optimizada, el 74.6% contaba con servicios de saneamiento mejorados y el 91.3% contaba con electricidad. La duración promedio de la educación materna fue de 7.8 años (DE: 4.1), lo cual corresponde a una educación secundaria incompleta.

La Tabla I expone las propiedades de la muestra en función del estado nutricional. Los niños diagnosticados con stunting mostraron una mayor proporción de residir en residencias rurales (72.3% contra 62.6%, $p<0.001$), presentaron un acceso reducido a servicios de saneamiento mejorados (67.8% contra 76.2%, $p<0.001$), y presentaron madres con un nivel educativo inferior (6.4 contra 8.1 años, $p<0.001$). La prevalencia de morbilidad en el reciente estudio mostró similitudes entre distintos grupos.

TABLA I
CARACTERÍSTICAS DE LOS NIÑOS DE 0-59 MESES SEGÚN ESTADO NUTRICIONAL

Característica	Sin stunting (n=7,067)	Con stunting (n=1,646)	p-valor
Edad (meses), media (DE)	28.7 (17.1)	32.4 (17.5)	<0.001
Sexo masculino, n (%)	3,546 (50.2%)	892 (54.2%)	0.003
Área rural, n (%)	4,426 (62.6%)	1,190 (72.3%)	<0.001
Agua mejorada, n (%)	6,325 (89.5%)	1,446 (87.8%)	0.046
Saneamiento mejorado, n (%)	5,385 (76.2%)	1,116 (67.8%)	<0.001
Electricidad, n (%)	6,543 (92.6%)	1,406 (85.4%)	<0.001
Educación materna (años), media (DE)	8.1 (4.0)	6.4 (4.1)	<0.001

DE: Desviación estándar. Prueba t de Student para variables continuas; chi-cuadrado para variables categóricas

B. Prevalencia según determinantes socioeconómicos

Se registró un gradiente socioeconómico pronunciado en la prevalencia de la desnutrición crónica (Fig. 1). El quintil más desfavorecido (Q1) exhibió una prevalencia del 39.2%, superando en más de cinco veces al quintil más acaudalado (Q5, 7.1%). La prevalencia experimentó una reducción monótona a través de los quintiles intermedios: Q2, Q3 y Q4, respectivamente. Esta correlación dosis-respuesta indica un impacto causal del estrato socioeconómico en el estado nutricional de la población infantil.

La discrepancia entre áreas rurales y urbanas fue igualmente evidente: los niños de zonas rurales exhibieron una prevalencia del 23.4% en comparación con el 10.9% en zonas urbanas (razón de prevalencia: 2.15). La evaluación basada en la educación materna evidenció un gradiente protector

análogo: las madres sin educación formal presentaron hijos con una prevalencia del 30.1%, mientras que las madres con educación superior exhibieron una prevalencia del 9.9% en sus hijos.

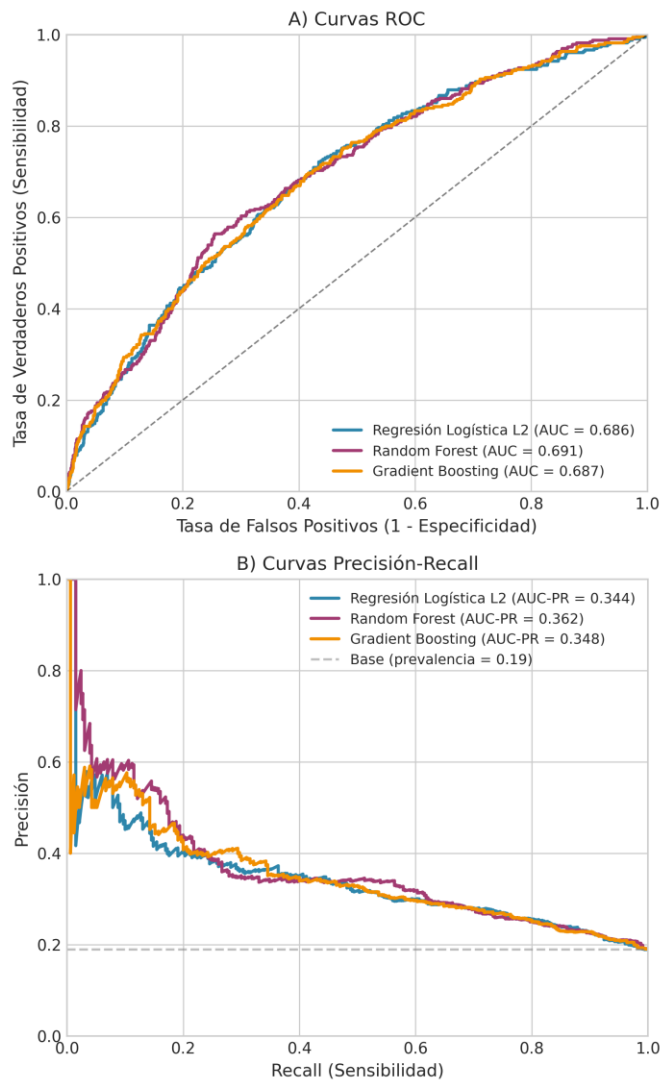


Fig. 1 Curvas ROC (izquierda) y Precisión-Recall (derecha) para los tres modelos de clasificación evaluados.

C. Desempeño de los modelos predictivos

La Tabla II sintetiza las métricas de rendimiento correspondientes a los tres modelos evaluados en el conjunto de prueba. El modelo Random Forest demostró un rendimiento discriminatorio superior con una AUC-ROC de 0.691 (Intervalo de Confianza del 95%: 0.668-0.714) y una AUC-PR de 0.362. El modelo Gradient Boosting exhibió un rendimiento análogo (AUC-ROC: 0.687), aunque exhibió una calibración superior (Brier Score: 0.145 contra 0.212). La regresión logística exhibió un rendimiento marginalmente inferior (AUC-ROC: 0.686), a pesar de mantener una mayor sensibilidad (0.721).

TABLA II
CARACTERÍSTICAS DE LOS NIÑOS DE 0-59 MESES SEGÚN ESTADO NUTRICIONAL

Modelo	AUC-ROC	AUC-PR	Brier	Sens	Espec.	F1
Regresión Logística L2	0.686	0.344	0.224	0.721	0.566	0.395
Random Forest	0.691	0.362	0.212	0.564	0.745	0.411
Gradient Boosting	0.687	0.348	0.145	0.697	0.584	0.220

Nota: AUC-ROC: Área bajo la curva ROC; AUC-PR: Área bajo la curva Precisión-Recall; Sens: Sensibilidad; Espec: Especificidad

D. Importancia de variables mediante SHAP

El análisis SHAP aplicado al modelo Gradient Boosting (Fig. 2) identificó el índice de riqueza como el predictor más influyente, registrando un valor SHAP medio absoluto de 0.489. Este valor fue seguido por la edad del niño (0.323) y la educación materna en años (0.251). Las variables asociadas con los servicios de agua, saneamiento y electricidad exhibieron aportaciones moderadas (rango: 0.06-0.12), mientras que la morbilidad reciente y las variables regionales exhibieron una relevancia relativamente reducida.

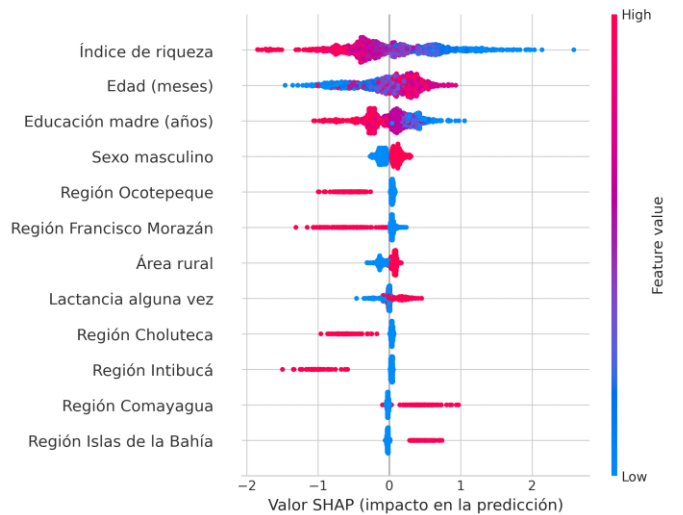


Fig. 2 Gráfico de resumen SHAP mostrando la contribución de cada variable a las predicciones. Cada punto representa una observación; el color indica el valor de la variable (azul=bajo, rojo=alto).

E. Relaciones funcionales y efectos de interacción

Los diagramas de dependencia SHAP (Fig. 3) desvelaron las interrelaciones funcionales entre los predictores y el riesgo de ocurrencia de stuning. El indicador de riqueza exhibió una correlación monotónica decreciente con valores SHAP negativos (protectores) que se iniciaron aproximadamente desde el percentil 40. La edad del infante exhibió una correlación no lineal positiva, con un incremento acelerado del riesgo entre los 12 y 36 meses, que posteriormente se estabilizó. La instrucción materna demostró rendimientos

marginales decrecientes: el efecto protector de cada año adicional fue más pronunciado durante los primeros 6-8 años de educación.

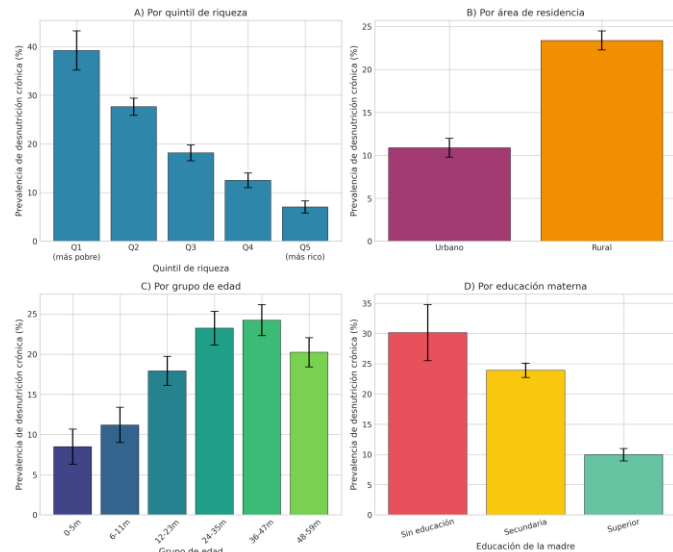


Fig. 3 Prevalencia de stunting según: (a) quintil de riqueza, (b) área de residencia, (c) grupo de edad, (d) educación materna.

El examen de la interacción evidenció un efecto sinérgico entre la pobreza y la ruralidad. En el quintil más desfavorecido (Q1), la prevalencia de desnutrición crónica fue del 42.8% en áreas rurales en contraposición al 31.4% en áreas urbanas, lo que representa una discrepancia absoluta de 11.4%. La discrepancia entre las regiones rural y urbana se disminuyó de manera progresiva en los quintiles superiores hasta casi desaparecer en Q5 (7.8% rural contra 6.5% urbano, diferencia: 1.3 puntos). Este patrón indica que la ubicación rural potencia el impacto deletéreo de la pobreza en el estado nutricional de la población infantil.

La Fig. 4 presenta los diagramas de dependencia SHAP que representan la contribución marginal de cuatro determinantes al riesgo estimado de desnutrición crónica por el modelo (valores SHAP > 0 incrementan el riesgo, valores SHAP < 0 lo disminuyen). Se evidencia una correlación inversa monotónica entre el índice de riqueza y el valor SHAP: los hogares con menor riqueza manifiestan contribuciones positivas (un incremento en el riesgo), mientras que los quintiles superiores exhiben contribuciones de manera consistente negativas (un efecto protector). Se observa una concentración de observaciones rurales en los niveles de riqueza más bajos, lo que sugiere una amplificación del riesgo en dichos contextos.

En (B), la edad del infante exhibe un patrón no lineal: el riesgo se incrementa de manera progresiva desde los primeros meses hasta cerca de la mitad de la franja de 0 a 59 meses (con máximos aproximados en 35 a 45 meses), y tiende a estabilizarse o disminuir levemente en edades avanzadas. Adicionalmente, la estratificación por riqueza sugiere que los valores más bajos del índice se correlacionan con

contribuciones SHAP más elevadas a lo largo del intervalo etario, lo que evidencia una interacción socioeconómica. En (C), se observa un gradiente protector evidente en la educación materna: con un incremento en la escolaridad, los valores SHAP descienden hacia valores negativos, con una disminución más pronunciada en niveles educativos medios-altos, y nuevamente con modulación por riqueza. En última instancia, (D) sintetiza el impacto del área de residencia: la distribución de SHAP es predominantemente positiva en áreas rurales y negativa en áreas urbanas, lo que sugiere que la residencia en áreas rurales funciona como un factor de riesgo autónomo en el modelo, en consonancia con las desigualdades territoriales detectadas.

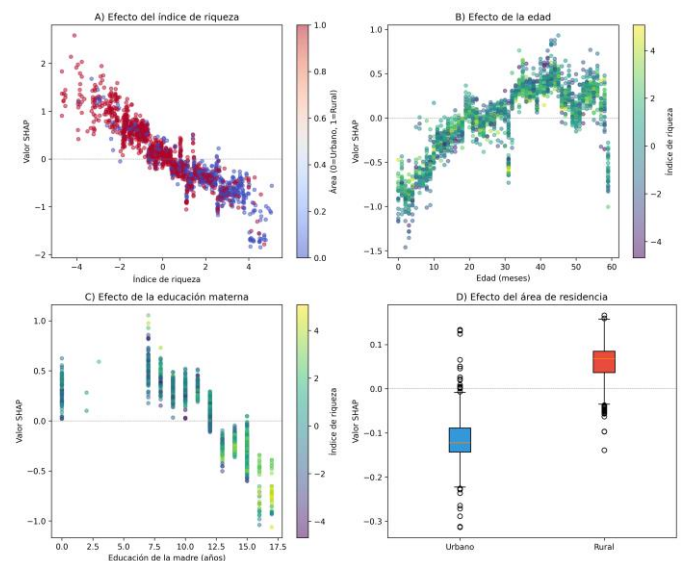


Fig. 4 Gráficos de dependencia SHAP para las cuatro variables principales. El eje X muestra el valor de la variable; el eje Y muestra su contribución al riesgo predicho.

En la Fig. 5 se muestra un mapa térmico de la prevalencia de desnutrición crónica (talla baja), estratificada simultáneamente por quintil de riqueza (Q1-Q5) y área de residencia (urbana/rural), revelando un gradiente socioeconómico y desigualdad espacial. En ambos contextos, la incidencia se reduce uniformemente a medida que aumenta la riqueza, desde un 30.0% en el primer quintil urbano y un 39.9% en el primer quintil rural hasta un 6.4% y un 8.5% en el quinto quintil urbano y rural, respectivamente. De forma sistemática, la residencia rural se asocia a prevalencias más altas que la urbana (diferencias absolutas: +9.9 en Q1; +11.2 en Q2; +7.0 en Q3; +2.7 en Q4; +2.1 en Q5 puntos porcentuales), lo que apunta a un efecto contextual adicional del territorio más allá de su posición socioeconómica. En resumen, el patrón indica que existe una asociación entre riqueza y ruralidad, donde la brecha urbano-rural se amplía en los quintiles más bajos y se reduce a medida que se avanza en los quintiles superiores, lo que apoya la focalización de intervenciones en hogares más pobres en zonas rurales.

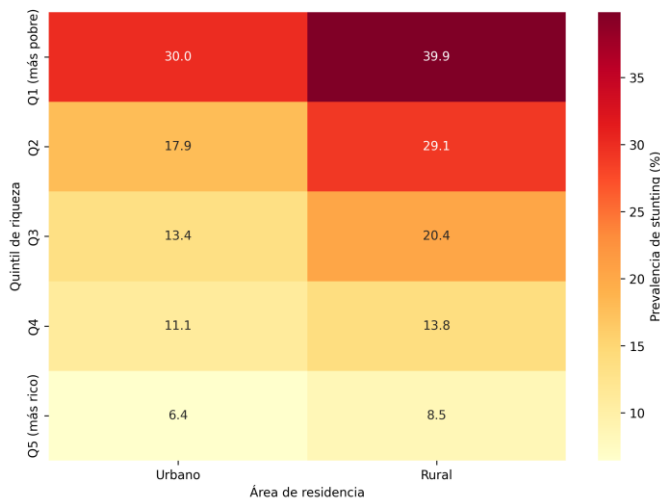


Fig. 5 Prevalencia de stunting (%) según quintil de riqueza y área de residencia, mostrando la interacción entre determinantes socioeconómicos y geográficos.

IV. DISCUSIÓN

Esta es la primera aplicación de aprendizaje automático explicable para analizar determinantes de desnutrición crónica infantil en Honduras. Los hallazgos confirman y miden la importancia de los determinantes estructurales socioeconómicos, revelando patrones de interacción no informados previamente con tal rigor metodológico.

La prevalencia de talla baja (18.9%) es comparable a las estimaciones del Informe Mundial sobre Nutrición 2022 para Honduras (18.7%), lo que confirma la representatividad de nuestra muestra analítica [24]. El gradiente socioeconómico observado, con una razón de prevalencia de 5.5 entre el quintil más pobre y el más rico, supera lo informado en otros países de la región como Guatemala (razón: 4.2) [25] y Perú (razón: 3.8) [26], lo que indica que las desigualdades socioeconómicas son especialmente marcadas en Honduras.

La capacidad predictiva de nuestros modelos (AUC-ROC: 0.686-0.691) es comparable a estudios similares que usaron datos de encuestas de Demografía y Salud. Un metanálisis de 2023 que incluyó 34 países reportó un AUC promedio de 0.689 para la predicción de stunting a partir de este tipo de encuestas [26], mientras que Bao et al. [27] obtuvieron una AUC de 0.71 en Bangladesh y Zhang et al. [11] informaron 0.68 en Papúa Nueva Guinea. Este techo de desempeño refleja una limitación estructural de los datos disponibles, no del enfoque de modelado. Varios determinantes con plausibilidad biológica establecida, incluyendo puntajes de diversidad dietética, calidad de la alimentación complementaria, estado nutricional materno prenatal, duración de la lactancia, biomarcadores de inflamación intestinal y contribuciones hereditarias al crecimiento lineal; no son capturados en el instrumento de la ENDESA/MICS 2019. Estudios que han integrado evaluación dietética con datos tipo EDS han demostrado ganancias incrementales en la capacidad

discriminativa de 3 a 8 puntos de AUC [28], [29], lo que sugiere que los modelos actuales se aproximan al techo práctico de predicción para este tipo de datos. Es fundamental distinguir entre dos propósitos analíticos: el tamizaje individual, para el cual un AUC inferior a 0.70 sería insuficiente, y la identificación y cuantificación de determinantes poblacionales modificables para orientar políticas públicas, propósito para el cual la interpretabilidad basada en SHAP es especialmente adecuada y para el cual el desempeño obtenido es razonablemente satisfactorio.

El hallazgo del índice de riqueza como el mejor predictor concuerda con el marco conceptual de UNICEF, que plantea la causalidad de la pobreza como la principal determinante del ciclo de desnutrición [7]. Este hallazgo tiene implicaciones directas para las políticas públicas: las intervenciones nutricionales aisladas, sin tener en cuenta las condiciones materiales de vida de los hogares, tendrán un efecto limitado. Programas de transferencias condicionadas, como el Bono Vida Mejor en Honduras, han mostrado impactos pequeños en indicadores nutricionales [30], lo que indica que las cantidades transferidas pueden ser insuficientes para cambiar sustancialmente las capacidades de consumo de alimentos en los hogares más vulnerables.

La asociación no lineal entre la edad infantil y el riesgo de talla baja, que se intensifica entre los 12 y 36 meses, corresponde con la ventana crítica para la alimentación complementaria y la exposición a patógenos entéricos. Este patrón ha sido ampliamente descrito en la literatura [31] y refleja el paso de la protección de la lactancia materna exclusiva a una mayor dependencia de la alimentación casera y la exposición ambiental. Las intervenciones dirigidas a esta etapa de la vida, como el asesoramiento nutricional de alta intensidad y la suplementación con micronutrientes, podrían maximizar el efecto sobre el crecimiento lineal.

La efectividad protectora de la educación materna, con rendimientos marginales decrecientes después de 6-8 años de escolaridad, indica que incluso alfabetizaciones básicas pueden modificar las prácticas de cuidado y nutrición infantil. Este hallazgo es consistente con la teoría del capital humano y la evidencia empírica de diferentes contextos [32]. La educación de las madres se expresa a través de varios mecanismos: una mayor comprensión de los mensajes de salud, mayor autonomía en las decisiones del hogar, mejores prácticas de higiene y mayor capacidad de generar ingresos.

La asociación que históricamente se ha encontrado entre pobreza y ruralidad tiene grandes implicaciones para focalizar las intervenciones. El aumento del efecto de la pobreza en zonas rurales se puede explicar porque se ven afectados varios mecanismos, como la disminución en la disponibilidad y variedad de alimentos, el acceso más lejano a servicios de salud, la mala calidad del agua y saneamiento, y la menor exposición a programas de desarrollo social [33]. El análisis SHAP proporciona una base territorial concreta para la segmentación geográfica. Entre las variables regionales incluidas en el modelo, Ocotepeque e Intibucá, ambos situados en el corredor occidental de Honduras emergieron como los

predictores regionales con los valores absolutos medios más altos de SHAP, indicando contribuciones residuales al riesgo que persisten incluso después de controlar el índice de riqueza, la educación materna y la residencia rural. Este hallazgo se ve reforzado por el mapa de calor de la pobreza-residencia (Fig. 5), que demuestra que los gradientes rural-urbano y de riqueza más pronunciados se concentran en los quintiles inferiores, el perfil más prevalente en los departamentos de las tierras altas occidentales. En conjunto, la señal cuantitativa de SHAP de estos departamentos específicos, combinada con la amplificación documentada de los efectos de la pobreza en entornos rurales, constituye la base empírica para recomendar que se prioricen paquetes de intervención de mayor intensidad en el corredor occidental. Este enfoque de segmentación geográfica es coherente con los marcos basados en la equidad que abogan por una asignación inversamente proporcional de recursos en relación con la necesidad [34], [35]. Esto sugiere que las estrategias de intervención deben ser territorialmente focalizadas, graduando paquetes de intervención más intensivos para poblaciones rurales pobres.

Las variables asociadas al agua y saneamiento mostraron aportes modestos pero relevantes al modelo predictivo. La evidencia actual ha confirmado la participación de la enteropatía ambiental, una inflamación crónica subclínica del intestino, causada por la exposición repetida a patógenos fecales, como un mediador entre la mala salud y el retraso en el crecimiento [36]. Las intervenciones holísticas que combinan mejoras en agua, saneamiento e higiene (WASH) con intervenciones nutricionales han demostrado tener efectos sinérgicos en el crecimiento lineal infantil [37].

La investigación muestra varias fortalezas metodológicas. La realización de una encuesta nacional representativa con mediciones antropométricas estandarizadas garantiza la validez externa de los resultados. La construcción del índice de riqueza usando la técnica de Análisis de Costos (ACP) evita los sesgos que implica medir directamente el ingreso en economías informales. El uso de algoritmos de aprendizaje automático permite verificar la robustez de los hallazgos ante distintas especificaciones. Los análisis SHAP finalmente abren la caja negra con una explicación teóricamente justificada.

Entre las limitaciones de la encuesta se encuentra su diseño transversal, que impide establecer relaciones de causa y efecto. La duración de la exposición a factores determinantes como la pobreza o la disponibilidad de servicios no se puede definir, y puede existir sesgo de causalidad inversa. Además, el cuestionario no incluye información sobre la diversidad dietética del niño, las prácticas de alimentación complementaria ni el estado nutricional materno, factores que podrían mejorar la capacidad predictiva de los modelos. El desempeño promedio de los modelos ($AUC < 0.70$) limita su uso para el cribado individual, pero su mayor potencial reside en identificar factores poblacionales para el desarrollo de políticas.

El aprendizaje automático interpretable puede revelar los factores que más contribuyen a la desnutrición crónica infantil en Honduras con mayor exactitud que los métodos estadísticos convencionales. Los análisis SHAP muestran que el nivel socioeconómico familiar es el mejor predictor, seguido de la edad del niño y la educación materna, con interacciones significativas entre pobreza y ruralidad.

Las desigualdades registradas, siendo cinco veces mayor en el quintil más pobre que en el más rico, demuestran que la desnutrición crónica en Honduras es, sobre todo, un problema de justicia social. Las intervenciones sólo nutricionales, sin modificar los determinantes estructurales, serán insuficientes y pueden aumentar las desigualdades si benefician principalmente a las poblaciones menos vulnerables.

Los resultados recomiendan cinco líneas prioritarias de acción para las políticas de nutrición infantil en Honduras: fortalecer y ampliar las transferencias monetarias focalizadas en los primeros 1,000 días de vida; acelerar las inversiones en agua y saneamiento rural, especialmente en el corredor occidental; implementar programas intensivos de educación nutricional durante la ventana de alimentación complementaria (6-24 meses); promover la educación de niñas y mujeres como estrategia de desarrollo intergeneracional; y usar enfoques de focalización geográfica que tomen en cuenta las interacciones entre pobreza y ruralidad.

Futuras investigaciones deberían incluir datos longitudinales para examinar trayectorias de crecimiento y establecer inferencias causales más fuertes. La combinación de datos sobre diversidad dietética, hábitos alimentarios y biomarcadores de inflamación intestinal podría mejorar en gran medida la capacidad predictiva de los modelos. Finalmente, estudios de implementación que evalúen la efectividad de intervenciones basadas en estos resultados serán necesarios para transformar la evidencia en mejoras reales en el estado nutricional de la niñez hondureña.

AGRADECIMIENTO

Agradecemos el apoyo recibido del Instituto de Investigación en Ciencias Médicas y Derecho a la Salud (ICIMEDS) 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). Agradecemos de a los Profesores Marcio Madrid e Isaac Zablah por su apoyo metodológico en este sentido. Los autores agradecen al profesor Isaac Zablah por su experta orientación metodológica y técnica, en múltiples aspectos de este trabajo.

REFERENCIAS

- [1] WHO Multicentre Growth Reference Study Group, "WHO Child Growth Standards: Length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age", World Health Organization, Geneva, 2006. [Online]. Available: <https://www.who.int/publications/i/item/924154693X>

- [2] R. E. Black et al., "Maternal and child undernutrition and overweight in low-income and middle-income countries", *The Lancet*, vol. 382, no. 9890, pp. 427-451, 2013. doi: 10.1016/S0140-6736(13)60937-X
- [3] UNICEF, WHO, and World Bank Group, *Levels and trends in child malnutrition: UNICEF/WHO/World Bank Group joint child malnutrition estimates: Key findings of the 2023 edition*. New York, NY, USA: UNICEF/WHO/World Bank, 2023.
- [4] Development Initiatives, "2022 Global Nutrition Report: Stronger commitments for greater action", Development Initiatives Poverty Research Ltd., Bristol, UK, 2022. [Online]. Available: https://globalnutritionreport.org/documents/922/2022_Global_Nutrition_Report.pdf
- [5] C. G. Victora et al., "Maternal and child undernutrition: consequences for adult health and human capital", *The Lancet*, vol. 371, no. 9609, pp. 340-357, 2008. doi: 10.1016/S0140-6736(07)61692-4
- [6] H. Alderman, J. R. Behrman, and S. Puett, "Big numbers about small children: Estimating the economic benefits of addressing undernutrition", *World Bank Research Observer*, vol. 32, no. 1, pp. 107-125, 2017. doi: 10.1093/wbro/lkw003
- [7] UNICEF, "Conceptual Framework on Maternal and Child Nutrition", UNICEF, New York, EEUU, 2020. [Online]. Available: <https://www.unicef.org/media/113291/file/UNICEFConceptualFramework.pdf>
- [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, 2016. doi: 10.1056/NEJMp1606181
- [9] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions", in *Adv. Neural Inf. Process. Syst. 30 (NeurIPS 2017)*, 2017, pp. 4765-4774.
- [10] A. Talukder, H. M. Ahammed, M. T. R. Prodhan, J. M. Hossain, and M. M. J. Uddin, "Machine learning algorithms for predicting malnutrition among under-five children in Bangladesh", *Nutrition*, vol. 78, art. no. 110861, 2020. doi: 10.1016/j.nut.2020.110861
- [11] J. Zhang et al., "Machine learning prediction of childhood stunting in Papua New Guinea", *Nutr. J.*, vol. 22, no. 1, p. 45, 2023. doi: 10.1186/s12937-023-00876-9
- [12] O. A. Obaído et al., "An interpretable machine learning approach for childhood malnutrition prediction using demographic and health survey data", *Sci. Rep.*, vol. 14, p. 8125, 2024. doi: 10.1038/s41598-024-58706-5
- [13] Secretaría de Salud de Honduras (Subsecretaría de Riesgos Poblacionales, Depto. de Epidemiología), *Análisis de situación de salud según condiciones de vida (ASIS-SCV): Honduras 2000*. [Online]. Available: <https://www.bvs.hn/E/pdf/AnalisisSalud.pdf>
- [14] Instituto Nacional de Estadística, "Encuesta Nacional de Demografía y Salud/Encuesta de Indicadores Múltiples por Conglomerados (ENDESA/MICS 2019)", INE Honduras, Tegucigalpa, 2020. [Online]. Available: <https://ine.gob.hn/bases-de-datos/>
- [15] WHO, "WHO Anthro software for personal computers: Software for assessing growth and development of the world's children", World Health Organization, 2011. [Online]. Available: <https://www.who.int/tools/child-growth-standards/software>
- [16] S. O. Rutstein and K. Johnson, "The DHS Wealth Index", DHS Comparative Reports No. 6, ORC Macro, Calverton, MD, 2004. [Online]. Available: <https://www.dhsprogram.com/pubs/pdf/cr6/cr6.pdf>
- [17] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. New York: Springer, 2009.
- [18] T. Saito and M. Rehmsmeier, "The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets", *PLoS ONE*, vol. 10, no. 3, p. e0118432, 2015. doi: 10.1371/journal.pone.0118432
- [19] Python Software Foundation, "Python 3.11 documentation", [Online]. Available: <https://docs.python.org/3.11/>
- [20] W. McKinney, "Data structures for statistical computing in Python", in *Proc. 9th Python in Science Conf. (SciPy)*, 2010, pp. 56-61, doi: 10.25080/Majora-92bf1922-00a
- [21] F. Pedregosa et al., "Scikit-learn: Machine learning in Python", *J. Mach. Learn. Res.*, vol. 12, pp. 2825-2830, 2011.
- [22] J. D. Hunter, "Matplotlib: A 2D graphics environment", *Comput. Sci. Eng.*, vol. 9, no. 3, pp. 90-95, 2007, doi: 10.1109/MCSE.2007.55
- [23] S. M. Lundberg et al., "From local explanations to global understanding with explainable AI for trees", *Nat. Mach. Intell.*, vol. 2, pp. 56-67, 2020. doi: 10.1038/s42256-019-0138-9
- [24] Development Initiatives, "Global Nutrition Report: Country Nutrition Profiles - Honduras", 2022. [Online]. Available: <https://globalnutritionreport.org/>
- [25] MSPAS Guatemala, "VI Encuesta Nacional de Salud Materno Infantil ENSMI 2014-2015", Ministerio de Salud Pública y Asistencia Social, Guatemala, 2017. [Online]. Available: <https://www.dhsprogram.com/pubs/pdf/fr318/fr318.pdf>
- [26] INEI Perú, "Encuesta Demográfica y de Salud Familiar ENDES 2022", Instituto Nacional de Estadística e Informática, Lima, 2023. [Online]. Available: https://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib1898/libro.pdf
- [27] B. Rao, M. Rashid, M. G. Hasan, and G. Thunga, "Machine Learning in Predicting Child Malnutrition: A Meta-Analysis of Demographic and Health Surveys Data", *Int. J. Environ. Res. Public Health*, vol. 22, no. 3, art. no. 449, 2025, doi: 10.3390/ijerph22030449
- [28] J. H. Rah, N. Akhter, R. D. Semba, S. de Pee, M. W. Bloem, A. A. Campbell, R. Moench-Pfanner, K. Sun, J. Badham, and K. Kraemer, "Low dietary diversity is a predictor of child stunting in rural Bangladesh", *Eur. J. Clin. Nutr.*, vol. 64, no. 12, pp. 1393-1398, 2010, doi: 10.1038/ejcn.2010.171
- [29] R. M. Gol, S. Kheirouri, and M. Alizadeh, "Association of Dietary Diversity With Growth Outcomes in Infants and Children Aged Under 5 Years: A Systematic Review", *J. Nutr. Educ. Behav.*, vol. 54, no. 1, pp. 65-83, 2022, doi: 10.1016/j.jneb.2021.08.016
- [30] J. Hoddinott and L. Bassett, "Conditional cash transfer programs and nutrition in Latin America: Assessment of impacts and strategies for improvement", International Food Policy Research Institute, Washington DC, 2008. doi: 10.2139/ssrn.1305326
- [31] P. Christian et al., "Risk of childhood undernutrition related to small-for-gestational age and preterm birth in low- and middle-income countries", *Int. J. Epidemiol.*, vol. 42, no. 5, pp. 1340-1355, 2013. doi: 10.1093/ije/dyt109
- [32] S. Vollmer et al., "Association between economic growth and early childhood undernutrition: Evidence from 121 Demographic and Health Surveys from 36 low-income and middle-income countries", *Lancet Glob. Health*, vol. 2, no. 4, pp. e225-e234, 2014. doi: 10.1016/S2214-109X(14)70025-7
- [33] L. E. Smith et al., "The effect of WASH interventions on linear growth in low and middle-income countries: A systematic review and meta-analysis", *Trop. Med. Int. Health*, vol. 24, no. 12, pp. 1387-1405, 2019. doi: 10.1111/tmi.13311
- [34] United Nations Children's Fund (UNICEF), Country programme document: Honduras, E/ICEF/2022/P/L.10, Dec. 27, 2021.
- [35] C. G. Victora, J. P. Vaughan, F. C. Barros, A. C. Silva, and E. Tomasi, "Explaining trends in inequities: evidence from Brazilian child health studies", *Lancet*, vol. 356, no. 9235, pp. 1093-1098, 2000, doi: 10.1016/S0140-6736(00)02741-0
- [36] A. J. Prendergast and P. Kelly, "Enteropathies in the developing world: Neglected effects on global health", *Am. J. Trop. Med. Hyg.*, vol. 86, no. 5, pp. 756-763, 2012. doi: 10.4269/ajtmh.2012.11-0743
- [37] A. J. Pickering et al., "The WASH Benefits and SHINE trials: Interpretation of WASH intervention effects on linear growth and diarrhoea", *Lancet Glob. Health*, vol. 7, no. 8, pp. e1139-e1146, 2019. doi: 10.1016/S2214-109X(19)30268-2