

Determinants of Childhood Non-Vaccination in Honduras: Using Classification, Interpretability, and Prediction Models

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– Childhood immunization is a cost-effective public health intervention to reduce morbidity and mortality in children under five years of age; however, in low- and middle-income countries there are coverage gaps that limit their impact. In Honduras, these gaps are challenging, so analytical approaches are needed to detect populations at risk of incomplete schemes. The study sought to identify the determinants of not fully vaccinated in children aged 12 to 23 months, develop predictive classification models and explain the contribution of each variable with SHAP values. The ENDESA/MICS 2019 was analyzed, including 1,711 children aged 12–23 months. The outcome was defined as not fully vaccinated, without BCG, three doses of Polio, three doses of Pentavalent and one of MMR. Logistic regression, decision trees, Random Forest and Gradient Boosting models were trained and compared, evaluating performance with AUC-ROC and Brier Score, and interpretability was examined with SHAP. The prevalence of non-vaccination was 19.3%. Notable performance was seen on Gradient Boosting (AUC-ROC = 0.694; Brier Score = 0.138). According to SHAP, the main determinants were the age of the child ($|SHAP| = 0.41$), possession of vaccination card ($|SHAP| = 0.31$) and maternal education ($|SHAP| = 0.10$). The department of Gracias a Dios had the highest prevalence of non-vaccination (51.8%). The findings indicate that machine learning with SHAP allows for the precise identification and prioritization of factors associated with incomplete vaccination schedules in Honduran children aged 12 to 23 months.

Keywords– Child vaccination, Machine learning, SHAP, Honduras, ENDESA/MICS.

Determinantes de la No Vacunación Infantil en Honduras: Usando Modelos de Clasificación, Interpretabilidad y Predicción

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 inmunización infantil es una intervención de salud pública costo-efectiva para reducir la morbilidad y menores de cinco años; sin embargo, en países de ingresos bajos y medios hay brechas de cobertura que limitan su impacto. En Honduras, estas brechas son un reto, por lo que se necesitan enfoques analíticos para detectar poblaciones en riesgo de esquemas incompletos. El estudio buscó identificar los determinantes de la no vacunación completa en niños de 12 a 23 meses, desarrollar modelos predictivos de clasificación y explicar el aporte de cada variable con valores SHAP. Se analizó la ENDESA/MICS 2019, incluyendo 1,711 niños de 12–23 meses. El desenlace se definió como no vacunación completa, sin BCG, tres dosis de Polio, tres de Pentavalente y una de SRP. Se entrenaron y compararon modelos de regresión logística, árboles de decisión, Random Forest y Gradient Boosting, evaluando desempeño con AUC-ROC y Brier Score, y se examinó la interpretabilidad con SHAP. La prevalencia de no vacunación fue de 19.3%. El mejor desempeño fue de Gradient Boosting (AUC-ROC = 0.694; Brier Score = 0.138). Según SHAP, los principales determinantes fueron la edad del niño ($|SHAP| = 0.41$), la tenencia de carne de vacunación ($|SHAP| = 0.31$) y la educación materna ($|SHAP| = 0.10$). El departamento de Gracias a Dios tuvo la mayor prevalencia de no vacunación (51.8%). Los hallazgos indican que el aprendizaje automático con SHAP permite identificar y jerarquizar con precisión los factores asociados a esquemas incompletos de vacunación en niños hondureños de 12 a 23 meses.

Palabras clave— Vacunación infantil, Aprendizaje automático, SHAP, Honduras, ENDESA/MICS.

I. INTRODUCCIÓN

La inmunización infantil constituye una de las intervenciones de salud pública más costosas a escala global, previniendo anualmente entre 3.5 y 5 millones de fallecimientos debido a patologías como la difteria, el tétanos, la tosferina, el sarampión y la poliomielitis [1], pero muy efectiva a largo plazo. El objetivo de la Organización Mundial de la Salud (OMS) es que todos los países alcancen coberturas de inmunización que superen el 90% para las vacunas básicas del programa ampliado de inmunización [2]. No obstante, a pesar de los progresos notables en las décadas recientes, persisten significativas disparidades de cobertura que impactan de manera desproporcionada a las poblaciones vulnerables en naciones de ingresos bajos y medios [3].

Dentro de la región de las Américas, Honduras ha evidenciado avances significativos en la disminución de la

mortalidad infantil vinculada a patologías prevenibles mediante la inmunización. Sin embargo, se han observado variaciones en la cobertura de vacunas en años recientes, con una tendencia decreciente que se intensificó durante la pandemia de COVID-19 [4]. Esta circunstancia presenta un riesgo epidemiológico considerable, dado que la acumulación de individuos susceptibles puede propiciar brotes de patologías previamente controladas o en proceso de erradicación.

Los factores determinantes de la no inmunización infantil son multifacéticos y complejos, incorporando elementos individuales, familiares, comunitarios y del sistema sanitario [5]. La bibliografía ha reconocido de manera consistente variables como el nivel educativo materno, el estatus socioeconómico, la zona de residencia (urbana/rural), el acceso geográfico a los servicios sanitarios y las características del sistema sanitario como indicadores de la cobertura vacunal [6]–[8]. No obstante, gran parte de investigaciones anteriores han recurrido a análisis estadísticos convencionales que presuponen relaciones lineales y aditivas entre predictores, lo cual puede subestimar las interacciones complejas entre variables.

Los procedimientos de aprendizaje automático han evidenciado una habilidad superior para identificar patrones no lineales y efectos de interacción en datos epidemiológicos [9]. En concreto, los modelos fundamentados en árboles de decisión, tales como Random Forest y Gradient Boosting, han demostrado un rendimiento predictivo efectivo en problemas de clasificación en el campo de la salud pública [10]. Sin embargo, una restricción histórica de estos procedimientos ha sido su carácter de "caja negra", que obstaculiza la interpretación de las predicciones en términos clínicamente pertinentes.

Los valores SHAP (SHapley Additive exPlanations), derivados de la teoría de juegos cooperativos, ofrecen una metodología rigurosa para cuantificar la contribución individual y global de las predicciones de modelos de aprendizaje automático [11]. Esta metodología posibilita la integración del potencial predictivo de algoritmos complejos con la interpretación requerida para la formulación de políticas y programas sanitarios.

La finalidad de esta investigación es identificar los determinantes de la no vacunación completa en niños hondureños de 12 a 23 meses, empleando datos de la Encuesta Nacional de Demografía y Salud (ENDESA/MICS 2019), construir modelos predictivos a través de técnicas de clasificación supervisada y proporcionar interpretaciones accionables a través del análisis SHAP. Los descubrimientos tienen como objetivo contribuir a la concentración de intervenciones destinadas a mitigar las deficiencias en la cobertura de vacunas en Honduras.

II. MATERIALES Y MÉTODOS

A. Fuente de Datos

Se llevó a cabo un análisis secundario de la Encuesta Nacional de Demografía y Salud utilizando la metodología MICS (ENDESA/MICS 2019) [12], llevado a cabo por el Instituto Nacional de Estadística de Honduras (INE) con asistencia técnica del Fondo de las Naciones Unidas para la Infancia (UNICEF). La encuesta se realizó mediante un diseño muestral estratificado por conglomerados en dos fases, con representatividad nacional, por área de residencia (urbana/rural) y por los 18 departamentos del país, junto con los municipios del Distrito Central y San Pedro Sula. La labor de campo se llevó a cabo durante los meses de julio y diciembre de 2019, previo a la pandemia de COVID-19.

A. Población de Estudios

Se incorporaron niños y niñas de 12 a 23 meses, residentes habituales en los hogares encuestados, conforme al estándar metodológico de la Organización Mundial de la Salud para la evaluación de la cobertura de esquemas completos de inmunización [2]. Este intervalo etario facilita la evaluación de la implementación del programa básico de inmunización, que debería haber sido completado durante el primer año de vida. Dentro de la base de datos de 8,713 niños menores de cinco años, 1,711 cumplieron con los criterios de elegibilidad, correspondientes a los 12-23 meses de edad al momento de la encuesta.

B. Definición del Desenlace

La variable de resultado fue la inmunización total, definida como la falta de cualquiera de las siguientes vacunas: una dosis de Bacilo Calmette-Guérin (Bacilo Calmette-Guérin), tres dosis de vacuna antipoliomielítica, tres dosis de vacuna pentavalente (DPT-HepB-Hib) y una dosis de vacuna SRP (sarampión, rubéola, parotiditis). Este esquema fundamental es el sugerido por la Organización Mundial de la Salud y el Programa Amplio de Inmunización de Honduras [13].

El dato relativo a la inmunización se extrajo primordialmente del carné de vacunación, cuando este fue presentado y validado por el entrevistador. En situaciones en las que el carné no fue accesible, se empleó el informe materno relativo a las vacunas recibidas. Esta estructuración

de fuentes se adhiere a las directrices metodológicas establecidas para el análisis de encuestas como ser la Demografía y Salud junto con la Encuesta Múltiple por Conglomerados (DHS/MICS) [14].

C. Variables Explicativas

Las variables explicativas fueron escogidas basándose en el marco teórico de determinantes de la inmunización infantil propuesto por Rainey et. al. [5]. Se tomaron en cuenta diversos factores relativos al niño, incluyendo: género (masculino/femenino), edad en meses y estado nutricional (desnutrición crónica definida como un puntaje Z de talla para edad inferior a -2 desviaciones estándar). Respecto al hogar y la madre/cuidador, se evaluaron aspectos como: el nivel educativo materno (preescolar, primaria (completa o menos), secundaria, superior), la zona de residencia (urbana/rural) y la posesión de un carné de vacunación. Se tomó en cuenta la región de residencia, clasificada como "más desarrollada" (Cortés - San Pedro Sula, Francisco Morazán - Distrito Central) en comparación con la región "menos desarrollada" (resto de departamentos).

D. Análisis Estadístico

El análisis descriptivo incorporó frecuencias absolutas y relativas para variables categóricas, así como medidas de tendencia central y dispersión para variables continuamente variables. Se realizaron estimaciones de las prevalencias globales de la no vacunación completa y se estratificaron por subgrupos de interés. Las discrepancias entre los grupos fueron evaluadas a través de las pruebas de Chi-cuadrado de Pearson, considerando significativo un valor p que no excede de 5 [15], [16].

No se incorporaron pesos de encuesta, estratos e identificadores de grupos en los modelos analíticos. Para las estimaciones descriptivas de prevalencia, se realizaron análisis sobre la muestra analítica no ponderada; Los lectores deben interpretar estas cifras como estimaciones basadas en muestras en lugar de proyecciones a nivel poblacional corregidas por diseño. Para el componente de modelado predictivo, esta decisión es metodológicamente coherente con el objetivo principal del estudio, identificar factores asociados a la no vacunación dentro de la muestra en lugar de producir estimaciones nacionales ponderadas y se alinea con la práctica estándar en aplicaciones de aprendizaje automático utilizando datos de encuestas complejas [17]. Se reconoce el posible efecto de este enfoque sobre las estimaciones de prevalencia como una limitación.

Para el análisis predictivo se emplearon cuatro algoritmos de clasificación: la regresión logística con regularización, el árbol de decisión, el árbol aleatorio y el incremento de la gradiente. Los datos fueron aleatoriamente segmentados en el conjunto de entrenamiento (75%) y prueba (25%), seleccionando el desenlace para mantener proporciones equivalentes. La selección de hiperparámetros se realizó mediante la validación cruzada de cinco pliegos. Dado el desequilibrio de clases (19.3% de casos positivos), se

implementó la ponderación inversa de clases en los modelos que permiten tal procedimiento.

El rendimiento de los modelos fue evaluado a través de: el área bajo la curva ROC (AUC-ROC), el área bajo la curva precisión-recall (AUC-PR), la exactitud, la precisión, la sensibilidad, el puntaje F1 y el puntaje Brier. La evaluación visual de la calibración se realizó a través de curvas de calibración. Se realizó un análisis de la interpretabilidad utilizando valores SHAP calculados utilizando el algoritmo TreeExplainer para modelos fundamentados en árboles [17]. Se registraron importancias globales (media del valor absoluto SHAP) y gráficos de dependencia correspondientes a las variables más significativas.

La evaluación se llevó a cabo en Python 3.11, empleando las bibliotecas pandas, scikit-learn, shap y matplotlib para su implementación. La información correspondiente a la encuesta fue extraída del repositorio oficial del Instituto Nacional de Estadística de Honduras [18]-[21].

E. Consideraciones Éticas

La investigación empleó información secundaria anonimizada de acceso público. La ENDESA/MICS 2019, aprobada por el Comité de Ética del Instituto Nacional de Estadística de Honduras, ha obtenido su aprobación. Los microdatos se encuentran accesibles para la investigación conforme a los términos de uso establecidos por UNICEF y el gobierno hondureño.

III. RESULTADOS

A. Características de la Muestra

La muestra analítica comprendió 1,711 niños en el rango de edad de 12 a 23 meses. La distribución de género resultó balanceada (50.4% de hombres). El 63.8% de la población infantil residía en zonas rurales y en regiones categorizadas como "menos desarrolladas" (81.1%). En relación con la educación materna, el 56.3% de las madres contaban con educación primaria, el 32.8% secundaria y el 7.0% superior. El 90.6% de los niños contaba con acceso a un carné de inmunización. La desnutrición crónica incidía en el 18.5% de la población estudiada. La Tabla 1 expone las características integrales de la muestra seleccionada.

B. Prevalencia de No Vacunación

La cobertura de vacunas individuales se determinó como: BCG representa el 94.4%, tres dosis de Polio son el 90.5%, tres dosis de Pentavalente son el 92.7%, tres dosis de Neumococo son el 91.4%, dos dosis de Rotavirus son el 92.1% y una dosis de SRP 89.5%. La cobertura del esquema básico completo (BCG + Polio3 + Penta3 + SRP1) fue del 80.7%, lo que se traduce en una prevalencia de no vacunación completa del 19.3% (Intervalo de Confianza del 95%: 17.4%-21.2%). La proporción de infantes denominados "zero-dose" (sin la administración de ninguna vacuna) ascendió al 1.8% (Ver Fig. 1).

TABLA I
CARACTERÍSTICAS DE LA MUESTRA Y PREVALENCIA DE NO VACUNACIÓN
SEGÚN VARIABLES SELECCIONADAS (N=1,711)

Variable	n	% Total	No. Vac. (n)	% No Vac.
Sexo				
Masculino	863	50.4	163	18.9
Femenino	848	49.6	167	19.7
Área de residencia				
Urbano	619	36.2	115	18.6
Rural	1,092	63.8	215	19.7
Educación materna				
Primaria o menos	963	56.3	195	20.2
Secundaria	562	32.8	98	17.4
Superior	119	7.0	22	18.5
Tenencia de carné				
Si	1,550	90.6	249	16.1
No	161	9.4	81	50.3

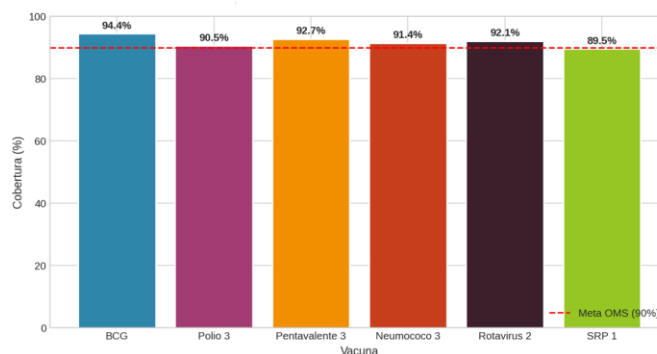


Fig. 1 Cobertura de vacunación en niños de 12-23 meses, Fuente: ENDESA/MICS 2019 [12]. Estimaciones muestrales no ponderadas; N = 1.711 niños de entre 12 y 23 meses.

El análisis por departamentos reveló marcadas disparidades geográficas. Gracias a Dios presentó la mayor prevalencia de no vacunación (51.8%), seguido de San Pedro Sula (35.7%) y Atlántida (35.5%). Por el contrario, Valle (6.6%), Lempira (8.2%) e Intibucá (9.3%) mostraron las prevalencias más bajas (Ver Fig. 2).

La tenencia de carné de vacunación mostró la asociación más fuerte con el desenlace. Los niños sin carné tuvieron una prevalencia de no vacunación de 50.3% comparado con 16.1% entre quienes lo tenían ($\chi^2 = 107.69$, $p < 0.001$). Las diferencias por sexo ($p = 0.718$), área de residencia ($p = 0.620$) y tipo de región ($p = 1.000$) no fueron estadísticamente significativas.

C. Desempeño de Modelos Predictivos

La Tabla 2 y Fig. 3 exponen las métricas de rendimiento correspondientes a los cuatro modelos examinados. El modelo Gradient Boosting demostró la mayor exactitud (82.2%) y precisión (63.0%), mientras que el modelo Random Forest exhibió el AUC-ROC más elevado (0.700).

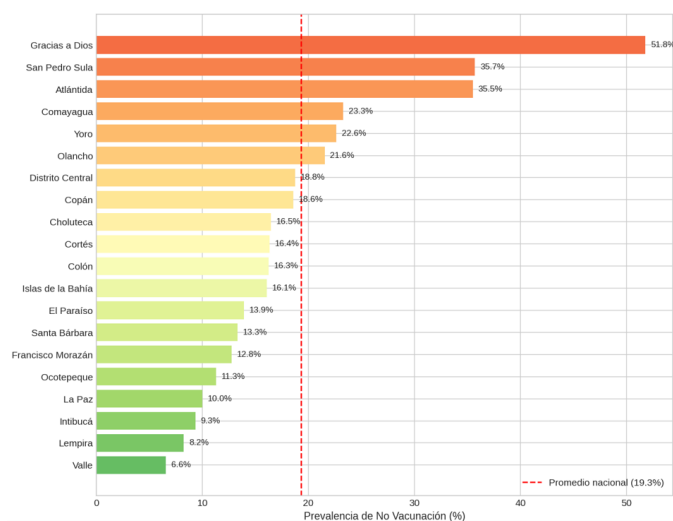


Fig. 2 Prevalencia de no vacunación completa por departamento. Estimaciones muestrales no ponderadas; N = 1.711 niños de entre 12 y 23 meses.

La regresión logística exhibió una sensibilidad superior (57.8%), aunque conllevaba una precisión reducida (29.1%). El Gradient Boosting logró la puntuación Brier más baja (0,138), frente a 0,204 para Random Forest, 0,210 para el Árbol de Decisión y 0,225 para Regresión Logística. La puntuación de Brier, interpretada como el error cuadrático medio de probabilidades predichas, penaliza tanto la mala discriminación como la mala calibración; Por tanto, valores más bajos indican una calibración de probabilidad superior.

TABLA II
DESEMPEÑO DE MODELOS DE CLASIFICACIÓN EN EL CONJUNTO DE PRUEBA

Modelo	Exac.	Prec.	Sen.	F1	AUC	Brier
Reg. Logística	0.645	0.291	0.578	0.387	0.674	0.225
Árbol Decisión	0.773	0.424	0.470	0.446	0.680	0.210
Random Forest	0.794	0.470	0.470	0.470	0.700	0.204
Gradient Boost.	0.822	0.630	0.205	0.309	0.694	0.138

Nota: Exac. = Exactitud; Prec. = Precisión; Sen. = Sensibilidad; AUC = Área bajo la curva ROC; Brier = Brier Score

Random Forest logró el AUC-ROC más alto (0,700), ligeramente por encima de Gradient Boosting (0,694). Dado que el objetivo analítico principal del estudio era producir estimaciones de riesgo individuales bien calibradas para apoyar intervenciones dirigidas, en lugar de maximizar solo el orden de rangos; se seleccionó Gradient Boosting como modelo preferido en función de su calibración de probabilidad superior (puntuación Brier más baja) junto con el rendimiento discriminativo competitivo.

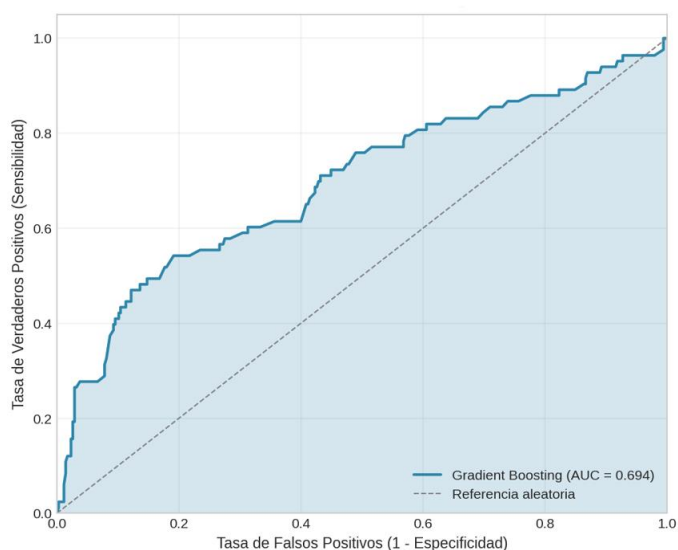


Fig. 3 Curva ROC del modelo Gradient Boosting.

D. Análisis de Interpretabilidad (SHAP)

El análisis SHAP del modelo Gradient Boosting determinó que la edad del niño es la variable con mayor contribución global a las predicciones ($|\text{SHAP}|$ medio = 0.41), seguida de la posesión de un carné de vacunación ($|\text{SHAP}| = 0.31$) y la educación materna ($|\text{SHAP}| = 0.10$). El área de residencia ($|\text{SHAP}| = 0.09$), la incidencia de desnutrición ($|\text{SHAP}| = 0.08$), y el tipo de región ($|\text{SHAP}| = 0.07$) exhibieron contribuciones inferiores (Fig. 4).

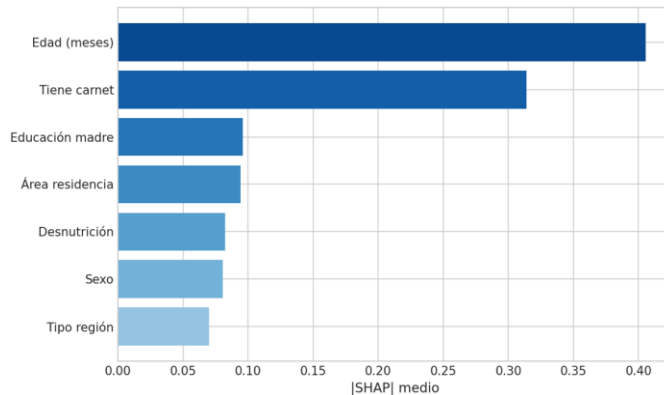


Fig. 4 Importancia de variables según valores SHAP medios absolutos

La relevancia significativa de la edad evidencia que los niños de 12 meses presentan una probabilidad inferior de recibir un esquema completo en comparación con los de 23 meses, lo cual se alinea con la dinámica temporal inherente al protocolo de inmunización. La posesión de un carné constituye el elemento modificable más significativo: su ausencia incrementa de manera significativa la probabilidad anticipada de no vacunación, lo que indica que el acceso y monitorización en los servicios de salud son factores determinantes esenciales.

IV. DISCUSIÓN

Este estudio detectó una prevalencia del 19.3% de no vacunación completa en niños hondureños de 12 a 23 meses, lo que constituye una discrepancia significativa en relación con los objetivos de cobertura universal. Este descubrimiento se alinea con informes anteriores provenientes de la región centroamericana y manifiesta retos continuos en los programas de inmunización [22], [23]. La implementación de técnicas de aprendizaje automático con interpretabilidad SHAP posibilitó la identificación del carné de vacunación como el factor modificable predominante asociado a la no vacunación, con consecuencias directas para el diseño de intervenciones.

Las notables discrepancias geográficas detectadas requieren un énfasis particular. De manera gracias a Dios, un departamento con una población predominantemente indígena (Misquitos, Tawahkas, Pech y Garífunas), con un acceso geográfico restringido y una densidad de servicios sanitarios reducida, registró una prevalencia de no vacunación tres veces superior al promedio nacional. Este modelo se alinea con el corpus literario que registra inequidades sistemáticas en el acceso a servicios de salud para las poblaciones indígenas en América [24]. Además, la prevalencia elevada en San Pedro Sula, la segunda urbe más poblada del país sugiere que los factores determinantes no son meramente geográficos, sino que también están vinculados con obstáculos sociales y organizativos del sistema sanitario en entornos urbanos complejos.

La falta de correlaciones significativas entre el sexo, la zona de residencia o la tipología de región en el análisis bivariado contrasta con los descubrimientos de otros contextos en los que estas variables desempeñan un papel predictivo significativo [25]. La ausencia de asociaciones estadísticamente significativas entre el área de residencia y la no vacunación en el análisis bivariado ($p = 0.620$) no contradice su contribución no nula en el modelo predictivo basado en SHAP ($|SHAP|$ media = 0.09). Estos dos hallazgos reflejan cantidades fundamentalmente diferentes. La prueba del qui-cuadrado mide la asociación marginal e incondicional entre un único predictor y el resultado, y puede carecer de potencia para detectar efectos modestos en presencia de un factor dominante de confusión, como la posesión de una tarjeta de vacunación. En cambio, los valores SHAP cuantifican la contribución incremental de cada variable a las predicciones del modelo tras tener en cuenta todas las demás características, una cantidad condicional y multivariante. Los modelos de aumento de gradiente pueden detectar además efectos de interacción que son invisibles para las pruebas por pares: por ejemplo, el área de residencia puede modificar la relación entre la posesión de la tarjeta y la no vacunación de formas que un chi-cuadrado bivariado no puede revelar. Además, la posesión de la tarjeta de vacunación puede actuar como mediador parcial de la vía relacionada con el acceso por la cual la residencia rural predeciría la no vacunación, atenuando su asociación marginal mientras preserva su señal predictiva condicional. En conjunto, estos hallazgos sugieren

que el área de residencia mantiene relevancia predictiva en un contexto multivariante, incluso cuando su asociación marginal no es detectable, un patrón documentado en aplicaciones epidemiológicas de aprendizaje automático [26]. El análisis SHAP corrobora que, si bien la educación materna y la zona de residencia contribuyen a las predicciones, su impacto es considerablemente inferior al de la posesión de carné.

La capacidad predictiva de los modelos (AUC-ROC de 0.67 a 0.70) es equiparable a investigaciones similares en naciones de ingresos bajos y medios [27]. La capacidad predictiva moderada evidencia que la no vacunación también está influenciada por factores no capturados en la encuesta, tales como la disponibilidad de vacunas en centros sanitarios, las prácticas de los proveedores y las posturas parentales respecto a la vacunación.

Se justifica una nota sobre el equilibrio entre sensibilidad y especificidad para quienes consideran aplicaciones operativas de estos modelos. Utilizando el umbral de clasificación por defecto de 0.5, el Gradient Boosting alcanzó la mayor precisión (63,0%) pero la menor sensibilidad (20,5%) entre los cuatro modelos, lo que significa que aproximadamente cuatro de cada cinco niños con vacunación incompleta no serían señalados por debajo de este umbral. Esto es consecuencia directa del desequilibrio de clases (19,3% de casos positivos) y de la optimización del modelo hacia la precisión global. Para una aplicación de cribado de salud pública, donde el objetivo principal es maximizar la detección de casos en lugar de minimizar falsos positivos; un umbral de clasificación más bajo o un modelo que priorice la sensibilidad, como la Regresión Logística (sensibilidad = 57,8%), puede ser operativamente preferible. Ajustar el umbral de decisión es un paso sencillo tras el entrenamiento que sacrifica precisión por recuerdo y puede calibrarse para ajustarse a la capacidad de intervención disponible. Los valores AUC-PR reportados junto con AUC-ROC ofrecen una imagen de rendimiento más informativa bajo el desequilibrio de clase, ya que la curva de precisión-recuerdo refleja directamente el equilibrio entre sensibilidad y valor predictivo positivo en diferentes umbrales. Recomendamos que cualquier posible despliegue de estos modelos vaya precedido de un análisis umbral guiado por el coste de los casos perdidos en relación con los falsos positivos en el contexto programático específico.

La selección del modelo priorizó la calibración junto con la discriminación. Mientras que Random Forest logró un AUC-ROC ligeramente superior (0,700 frente a 0,694), Gradient Boosting demostró la mejor calibración de probabilidad en todos los modelos (Brier Score = 0,138), como lo confirmaron visualmente las curvas de calibración. Para aplicaciones centradas en la estimación de puntuaciones de riesgo donde la magnitud de las probabilidades predichas informa la asignación de recursos los modelos bien calibrados suelen ser preferibles a los que simplemente clasifican correctamente a los individuos. Sobre esta base, se seleccionó el Gradient Boosting como el modelo principal para el análisis de interpretabilidad SHAP.

Esta investigación posee fortalezas significativas. Emplea información representativa a escala nacional mediante una metodología estandarizada a nivel internacional. La integración de modelos predictivos con la interpretación SHAP facilita la superación de las restricciones inherentes a los enfoques puramente descriptivos o a los modelos de caja negra. No obstante, es imperativo tener en cuenta ciertas restricciones. Además, los pesos de muestreo de la encuesta no se aplicaron a los análisis estadísticos ni al de aprendizaje automático. Aunque este enfoque es apropiado para un objetivo de modelización predictiva centrado en la identificación de patrones dentro de la muestra, significa que las estimaciones de prevalencia puntual pueden diferir de las cifras poblacionales ponderadas por diseño. Los análisis destinados a producir estimaciones de prevalencia representativas a nivel nacional de esta encuesta deben incorporar los pesos muestrales ENDESA/MICS 2019 utilizando métodos complejos de encuesta. El diseño transversal obstaculiza la identificación de relaciones causales. La declaración materna de inmunización sin carné puede inducir sesgo de recuerdo. Las variables disponibles no incorporan indicadores directos del acceso a servicios, actitudes hacia la inmunización ni rasgos del sistema sanitario a escala local.

Los descubrimientos poseen relevancia para la estrategia sanitaria en Honduras. Inicialmente, las tácticas destinadas a optimizar la cobertura de vacunas deben dar prioridad a la distribución y registro sistemático de los carnés de vacunación. Este proceso no solo simplifica el seguimiento individual, sino que también funciona como un indicador de interacción con los servicios de salud. En segunda instancia, es imperativo implementar intervenciones focalizadas en los departamentos identificados como de alto riesgo, especialmente en Gracias a Dios, donde probablemente se requieran estrategias culturalmente adaptadas y logísticamente adecuadas para comunidades dispersas. En última instancia, la incorporación de herramientas predictivas fundamentadas en datos puede facilitar la detección anticipada de niños que podrían no haber completado esquemas de inmunización.

V. CONCLUSIÓN

La combinación de modelos predictivos y métodos de interpretabilidad aporta evidencia clara sobre las variables que más contribuyen a las brechas de cobertura, ofreciendo insumos útiles para orientar intervenciones y fortalecer las estrategias nacionales de inmunización. Estos hallazgos subrayan el valor de incorporar enfoques analíticos avanzados en la toma de decisiones en salud pública, especialmente en contextos donde persisten desigualdades en el acceso a servicios preventivos. Se recomienda fortalecer los sistemas de registro de vacunación y desarrollar estrategias específicas para las regiones con mayor brecha de cobertura.

AGRADECIMIENTO

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). Los autores agradecen a los profesores Marcio Madrid e Isaac Zablah por su experta orientación metodológica y técnica, en múltiples aspectos de este trabajo.

REFERENCIAS

- [1] World Health Organization, "Immunization coverage", *WHO Fact Sheets*, 2023. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/immunization-coverage>
- [2] World Health Organization, *Immunization Agenda 2030: A Global Strategy to Leave No One Behind*. Geneva, Switzerland: WHO, 2020.
- [3] M. Peck, M. Gacic-Dobo, M. S. Diallo, Y. Nedelec, S. S. Sodha, and A. S. Wallace, "Global routine vaccination coverage, 2018", *MMWR Morb. Mortal. Wkly. Rep.*, vol. 68, no. 42, pp. 937–942, 2019, doi: 10.15585/mmwr.mm6842a1
- [4] Pan American Health Organization, *Immunization in the Americas: 2022 Summary*. Washington, DC, USA: PAHO, 2022.
- [5] J. A. Rainey, M. Watkins, T. K. Ryman, R. W. Sandhu, A. Bo, and K. Banerjee, "Reasons related to non-vaccination and under-vaccination of children in low and middle income countries: Findings from a systematic review of the published literature, 1999–2009", *Vaccine*, vol. 29, no. 46, pp. 8215–8221, 2011, doi: 10.1016/j.vaccine.2011.08.096.
- [6] J. L. Mathew, "Inequity in childhood immunization in India: A systematic review", *Indian Pediatr.*, vol. 49, no. 3, pp. 203–223, Mar. 2012, doi: 10.1007/s13312-012-0063-z.
- [7] D. Lakew, A. Bekele, and S. Biadgilign, "Factors influencing full immunization coverage among 12–23 months of age children in Ethiopia: Evidence from the national demographic and health survey in 2011", *BMC Public Health*, vol. 15, Art. no. 728, 2015, doi: 10.1186/s12889-015-2078-6.
- [8] A. A. Adenike, O. I. Adejumo, E. T. Olufunmi, and O. J. Ridwan, "Maternal characteristics and immunization status of children in North Central of Nigeria", *Pan Afr. Med. J.*, vol. 26, Art. no. 159, 2017, doi: 10.11604/pamj.2017.26.159.11530
- [9] S. L. Goldstein, S. J. Ausiello, and E. J. Benjamin, "Machine learning in cardiovascular medicine: Are we there yet?", *Heart*, vol. 103, no. 8, pp. 563–565, 2017, doi: 10.1136/heartjnl-2016-310835
- [10] A. Rajkomar, J. Dean, and I. Kohane, "Machine learning in medicine", *N. Engl. J. Med.*, vol. 380, no. 14, pp. 1347–1358, 2019, doi: 10.1056/NEJMr1814259
- [11] 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
- [12] Instituto Nacional de Estadística de Honduras, *Encuesta Nacional de Demografía y Salud ENDESA/MICS 2019: Informe Final*. Tegucigalpa, Honduras: INE, 2020. [Online]. Available: <https://inec.gob.hn/endesa/>
- [13] Secretaría de Salud de Honduras, *Programa Ampliado de Inmunizaciones: Esquema Nacional de Vacunación 2019*. Tegucigalpa, Honduras: SESAL, 2019
- [14] ICF International, *The DHS Program: Guide to DHS Statistics*. Rockville, MD, USA: ICF International, 2018
- [15] L. Madrigal, "Descriptive statistics: measures of central tendency and dispersion", in *Statistics for Anthropology*. Cambridge, U.K.: Cambridge University Press, 1998, pp. 31–53, doi: 10.1017/CBO9781139171113.004
- [16] NIST/SEMATECH, "Critical Values of the Chi-Square Distribution", *NIST/SEMATECH e-Handbook of Statistical Methods*, [Online]. Available: <https://www.itl.nist.gov/div898/handbook/eda/section3/eda3674.htm>
- [17] T. Lumley, *Complex Surveys: A Guide to Analysis Using R*. Hoboken, NJ, USA: Wiley, 2011, doi: 10.1002/9780470580066.

- [18]Python Software Foundation, “Python 3.11 documentation”, [Online]. Available: <https://docs.python.org/3.11/>
- [19]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
- [20]F. Pedregosa *et al.*, “Scikit-learn: Machine learning in Python”, *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011
- [21]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
- [22]C. Restrepo-Méndez, A. J. Barros, A. M. Wong, H. L. Johnson, G. Pariyo, and C. G. Victora, “Inequalities in full immunization coverage: Trends in low- and middle-income countries”, *Bull. World Health Organ.*, vol. 94, no. 11, pp. 794–805B, 2016, doi: 10.2471/BLT.15.162172.
- [23]J. E. Ataguba, C. Ojo, and D. McIntyre, “Income-related health inequality in Sub-Saharan Africa: A decomposition analysis”, *Soc. Sci. Med.*, vol. 238, Art. no. 112460, 2019, doi: 10.1016/j.socscimed.2019.112460.
- [24]Pan American Health Organization, *Health of Indigenous Peoples in the Americas*. Washington, DC, USA: PAHO, 2016. [Online]. Available: <https://www.paho.org/sites/default/files/55-saludPI-decenio-Eng2.pdf>
- [25]N. Bbaale, “Factors influencing childhood immunization in Uganda”, *J. Health Popul. Nutr.*, vol. 31, no. 1, pp. 118–129, 2013, doi: 10.3329/jhpn.v31i1.14756.
- [26]G. Shmueli, “To explain or to predict?”, *Stat. Sci.*, vol. 25, no. 3, pp. 289–310, 2010, doi: 10.1214/10-STS330
- [27]K. Mosser, J. Rudhart, and S. I. Hay, “Mapping vaccination coverage to explore the effects of delivery mechanisms and inform vaccination policy”, *Nat. Commun.*, vol. 12, no. 1, Art. no. 1633, 2021, doi: 10.1038/s41467-021-21783-1.