Artificial Intelligence Based Child Nutritional Stratification System for Rural Communities: Classification Design and Community Surveillance Dashboard

Ethel Flores[1]; Johana Bermudez[2]; Diana Meza[3]; Jenny Cruz[4]; Yenny Velasquez[5]; Juan José Flores[6]
[1,2,3,4,5] Universidad Nacional Autónoma de Honduras, Honduras, *ethel.flores@unah.edu.hn, johana.bermudez@unah.edu.hn, magoe0101@gmail.com, jenykcm28@gmail.com, yenny.velasquez@unah.edu.hn*
[6] Hospital General San Felipe, Honduras, *juanjflores2003@yahoo.com*

Abstract– ***Child malnutrition remains a major public health and development challenge, especially in rural communities where continuous surveillance is difficult. This paper proposes an artificial intelligence based nutritional stratification system for children aged 6 to 59 months. The system combines rule based anthropometric safety criteria, a calibrated model that estimates the risk of nutritional deterioration within 60 days and a community dashboard for operational follow up. The architecture is specified using the C4 model and includes offline mobile capture, data quality control, z score computation, interpretable risk estimation, audit logs and alert management. A synthetic technical validation with 1,200 simulated children was used to verify workflow consistency, without claiming clinical effectiveness or epidemiological prevalence. Under the prespecified synthetic scenarios, the hybrid workflow combining safety rules with a calibrated XGBoost model showed the intended safety behavior, increasing the simulated capture of priority cases. These figures are implementation-test outputs and must not be interpreted as estimates of clinical predictive performance. The dashboard converts classification outputs into prioritized work lists, overdue visit alerts and territorial risk maps, enabling community health workers to allocate limited field resources according to explicit and auditable criteria. The main contribution is modelan integration and operational orchestration model, rather than a new learning algorithm, for converting community measurements into prioritized surveillance actions.***

Keywords-- ***artificial intelligence, child nutrition, community surveillance, digital health, machine learning.***

Sistema de estratificación nutricional infantil con inteligencia artificial para comunidades rurales: diseño de clasificación y tablero comunitario de vigilancia continua

Ethel Flores[1]; Johana Bermudez[2]; Diana Meza[3]; Jenny Cruz[4]; Yenny Velasquez[5]; Juan José Flores[6]
[1,2,3,4,5] Universidad Nacional Autónoma de Honduras, Honduras, *ethel.flores@unah.edu.hn, johana.bermudez@unah.edu.hn, magoe0101@gmail.com, jenykcm28@gmail.com, yenny.velasquez@unah.edu.hn*
[6] Hospital General San Felipe, Honduras, *juanjflores2003@yahoo.com*

***Resumen*– La desnutrición infantil continúa siendo un problema de salud pública y desarrollo humano, especialmente en comunidades rurales donde la vigilancia continua es limitada. Este artículo propone un sistema de estratificación nutricional infantil con inteligencia artificial para niños de 6 a 59 meses. El sistema combina reglas antropométricas de seguridad, un modelo calibrado que estima el riesgo de deterioro nutricional a 60 días y un tablero comunitario para seguimiento operativo. La arquitectura se especifica mediante el modelo C4 e incluye captura móvil offline, control de calidad de datos, cálculo de puntuaciones z, estimación interpretable de riesgo, bitácoras de auditoría y gestión de alertas. Se realizó una validación técnica 1,200 niños sintéticos para verificar consistencia del flujo, sin afirmar eficacia clínica ni estimar prevalencia epidemiológica. En los escenarios sintéticos preespecificados, el flujo híbrido de reglas de seguridad y XGBoost calibrado mostró el comportamiento de seguridad previsto, al aumentar la captura simulada de casos prioritarios. Estas cifras son resultados de prueba de implementación y no deben interpretarse como estimaciones de desempeño predictivo clínico. La principal contribución es un modelo de integración y orquestación operativa, no un nuevo algoritmo de aprendizaje, para convertir mediciones comunitarias en acciones priorizadas.**

***Palabras clave*-- inteligencia artificial, nutrición infantil, vigilancia comunitaria, salud digital, aprendizaje automático.**

I. Introducción

La desnutrición infantil continúa siendo un problema persistente de salud pública y desarrollo humano. Las estimaciones conjuntas UNICEF, OMS y Banco Mundial documentan que en 2024 había 150.2 millones de niños menores de cinco años con retraso del crecimiento, 42.8 millones con emaciación y 35.5 millones con sobrepeso, lo que confirma que la malnutrición debe entenderse como un continuo de déficit, exceso y vulnerabilidad alimentaria [1], [2]. En Honduras, los perfiles internacionales de nutrición reportan avances parciales, aunque el retraso del crecimiento en menores de cinco años se mantiene por encima del promedio regional, con una carga que afecta de forma más intensa a hogares rurales, dispersos y con menor acceso oportuno a servicios de salud [3].

La vigilancia nutricional comunitaria suele depender de rondas periódicas, instrumentos manuales y consolidación tardía de registros. En comunidades rurales, esta dinámica limita la capacidad de detectar deterioro entre visitas, priorizar desplazamientos domiciliarios y asignar recursos escasos. El problema no es únicamente estadístico, sino operativo: un equipo local puede tener datos de peso, talla o perímetro braquial medio, pero no siempre dispone de un mecanismo continuo que transforme esos datos en alertas verificables, listas de trabajo y seguimiento de acciones.

Este artículo propone un sistema de estratificación nutricional infantil con inteligencia artificial para comunidades rurales. El producto científico esperado es doble: un sistema de clasificación que combine reglas antropométricas basadas en estándares internacionales con un modelo interpretable de riesgo futuro, y un tablero comunitario que convierta la clasificación en una herramienta de vigilancia continua. La contribución tecnológica reside en integrar captura offline, control de calidad, cálculo antropométrico, priorización con aprendizaje automático, explicación local y gestión de alertas. La novedad que se reclama es de integración y orquestación operativa: cada componente está establecido por separado, y el aporte consiste en articularlos en un flujo auditable que distingue el estado clínico, el riesgo predicho y la prioridad operativa. La contribución de negocio público consiste en convertir datos comunitarios de bajo costo en una cadena de valor operativa para orientar visitas, referencia y supervisión territorial.

II. Antecedentes y fundamento técnico

Los estándares de crecimiento infantil de la OMS constituyen la base técnica para interpretar peso, longitud, talla e índice de masa corporal en menores de cinco años [4], [5]. Su uso permite calcular puntuaciones *z* comparables internacionalmente y clasificar condiciones como retraso del crecimiento, bajo peso o emaciación. La guía OMS de 2023 sobre emaciación y edema nutricional refuerza la necesidad de

identificar tempranamente la malnutrición aguda, reconocer edema bilateral y articular acciones de prevención, tratamiento y seguimiento [6]. Herramientas como WHO Anthro y los enfoques SMART han promovido estandarización antropométrica y control de plausibilidad, pero su orientación principal es analítica o de encuesta, no necesariamente de operación cotidiana comunitaria [7], [8].

La salud digital ofrece una oportunidad para mejorar este proceso cuando los sistemas se integran al flujo de trabajo real. La OMS ha recomendado intervenciones digitales bajo condiciones de aceptabilidad, factibilidad, equidad, capacidad de respuesta del sistema y protección de los usuarios [9]. La OMS también propuso una taxonomía de intervenciones digitales en salud que permite clasificar funcionalmente los sistemas según su propósito dentro del sistema de salud [10]. Para sistemas con inteligencia artificial, la OMS enfatiza supervisión humana, transparencia, responsabilidad, equidad y protección de datos [11]. En investigación de modelos predictivos, TRIPOD+AI establece requisitos de reporte para desarrollo, validación, actualización, uso de datos, desempeño y aplicabilidad de modelos clínicos basados en regresión o aprendizaje automático [12]. Por tanto, un clasificador nutricional no debe presentarse como sustituto del juicio clínico, sino como una capa de priorización auditable que preserve reglas de seguridad no anulables.

El aprendizaje automático puede mejorar la anticipación del deterioro cuando existen mediciones longitudinales, señales clínicas mínimas y variables de seguimiento. Modelos de árboles potenciados como XGBoost tienen buen desempeño en datos tabulares y permiten manejar interacciones no lineales [13]. No obstante, la interpretabilidad es indispensable en salud comunitaria. Métodos como SHAP permiten atribuir la contribución de cada variable a una predicción individual, lo que facilita explicar por qué un niño fue priorizado [14]. La evaluación debe combinar discriminación, calibración, utilidad clínica y seguridad, usando métricas como AUC ROC, Brier score, sensibilidad, especificidad, calibración y análisis de decisión cuando exista validación prospectiva [15]-[17].

III. Materiales y Métodos

Se empleó un enfoque de ciencia del diseño orientado a producir un artefacto tecnológico verificable. El sistema fue definido a partir de cuatro principios: mínima carga de captura, trazabilidad de cada clasificación, operación offline con sincronización diferida y supervisión humana en toda recomendación. El alcance se centró en niños de 6 a 59 meses atendidos por promotores comunitarios y personal de enfermería del primer nivel. Para menores de seis meses, el sistema solo registra mediciones y activa recomendación de evaluación clínica, debido a la complejidad de su valoración nutricional y a la necesidad de protocolos específicos.

La arquitectura fue expresada mediante el modelo C4, que permite describir un sistema a niveles de contexto, contenedores, componentes y código [18]. La Fig. 1 presenta el diagrama de contenedores. La aplicación móvil captura datos en campo y conserva una cola local cifrada cuando no hay conectividad. La API de sincronización valida identidad de dispositivo, recibe registros y envía actualizaciones. El motor antropométrico calcula puntuaciones z y banderas de plausibilidad. El modelo de riesgo estima la probabilidad de deterioro nutricional a 60 días. El módulo de explicabilidad genera razones locales de priorización y bitácoras de auditoría. La base de datos conserva registros codificados, y el tablero comunitario muestra alertas, cobertura, casos vencidos y distribución territorial de prioridades.

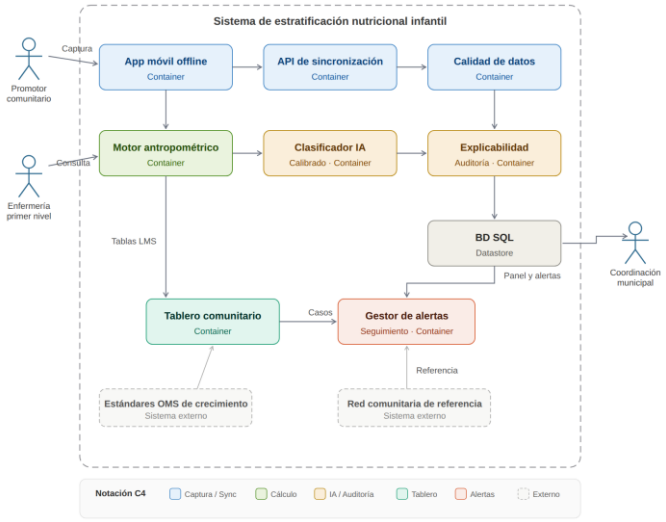


Fig. 1 Arquitectura del sistema expresada en notación C4 a nivel de contenedores. Se muestran la aplicación móvil de captura con cola local cifrada, la API de sincronización, el motor antropométrico, el modelo de riesgo de IA con módulo de explicabilidad, la base de datos codificada y el tablero comunitario. Las flechas indican dirección del flujo de datos.

El conjunto mínimo de datos se estructuró en cinco dominios: identificación codificada, antropometría, clínica mínima, trayectoria y contexto operativo. La identificación nominal no es necesaria para el análisis central y debe permanecer bajo custodia del equipo local. Las variables antropométricas incluyen edad en meses, sexo, peso, longitud o talla y perímetro braquial medio. La clínica mínima incluye edema bilateral, diarrea reciente, fiebre, apetito y signos de peligro. La trayectoria se deriva de visitas previas y captura cambios de peso, tendencia de puntuación z y visitas perdidas.

El contexto operativo incluye distancia o dificultad de acceso, disponibilidad alimentaria reportada y posibilidad de visita domiciliaria. Estas variables de contexto no se emplean como predictores del modelo de riesgo; solo alimentan el componente C_i del índice de prioridad definido en (2). La Tabla I resume este conjunto mínimo de datos y el rol de cada variable en el experimento sintético.

La clasificación opera en dos capas. La primera es una capa

de seguridad basada en reglas antropométricas y clínicas. Esta capa identifica referencia inmediata cuando existe edema bilateral, emaciación severa, perímetro braquial medio crítico, pérdida rápida de peso o signos de peligro. La segunda capa estima la probabilidad de deterioro nutricional en un horizonte operativo de 60 días. Para mantener transparencia, se entrenó una regresión logística calibrada como línea base y se compararon modelos de mayor capacidad, como bosque aleatorio y XGBoost.

Definición del resultado. El sistema distingue tres constructos que no deben confundirse: (a) el estado clínico actual, determinado por las reglas de la Tabla II; (b) el riesgo futuro predicho, p_i, y (c) la prioridad operativa, expresada por el índice I_i y el nivel de color. El origen temporal t_0 es la visita índice de cada niño, en la que se dispone de al menos dos mediciones previas. El resultado binario Y_i toma el valor 1 si, en alguna medición realizada en el intervalo (t_0, t_0 + 60 días], ocurre al menos una de estas condiciones: WHZ < −2 en un niño con WHZ ≥ −2 en t_0 (emaciación incidente), descenso de WHZ ≥ 0.5 respecto de t_0, MUAC < 12.5 cm en un niño con MUAC ≥ 12.5 cm en t_0, o aparición de edema bilateral. Todos los componentes se observan solo durante el seguimiento y ninguno se utiliza como predictor. En el escenario sintético el resultado se generó para todos los niños dentro del horizonte, sin censura; en datos reales, los niños sin medición en la ventana deberán tratarse como resultado desconocido y no como negativos.

Para la evaluación comparativa se empleó, además, un criterio operativo de caso prioritario: $Y_i = 1$ o presencia en t_0 de un criterio rojo o naranja de la capa de reglas. Este criterio compuesto refleja la función del sistema, que debe captar tanto el deterioro futuro como la gravedad ya presente. Por construcción, la capa de reglas detecta la segunda componente, lo que explica parte de la ganancia de sensibilidad del flujo híbrido y obliga a leerla como verificación de diseño, no como mejora predictiva.

TABLA I
VARIABLES MÍNIMAS, ROL EN EL EXPERIMENTO SINTÉTICO Y CONTROLES DE CALIDAD DEL SISTEMA

Grupo	**Variable principal**	**Uso**	**Control**	**Rol en el experimento**
Identificación	ID codificado, comunidad, fecha	Seguimiento longitudinal	Unicidad y trazabilidad	No predictor
Antropometría	Edad, sexo, peso, talla, MUAC	Puntuaciones z y reglas	Rangos plausibles	Entrada en t_0; WHZ y MUAC también en A_i
Clínica mínima	Edema, diarrea, fiebre, apetito	Referencia inmediata	Campos obligatorios	Entrada en t_0 (diarrea, fiebre, apetito); edema y peligro solo en reglas
Trayectoria	Cambio de peso, visitas perdidas	Riesgo anticipado	Mínimo dos visitas	Entrada en t_0; pendiente de WHZ también en T_i
Contexto	Distancia, acceso, disponibilidad	Priorización operativa	No excluir atención	Solo en C_i; no predictor
Seguimiento	WHZ, MUAC y edema en (t_0, t_0 + 60 d]	Definición de Y_i	Solo después de t_0	Resultado; nunca predictor

En un entorno con registros suficientes, el modelo de riesgo estima la probabilidad individual p_i según (1):

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

En (1), p_i representa la probabilidad de deterioro ($Y_i = 1$), x_{ij} son predictores estandarizados disponibles en t_0 y β_j son coeficientes del modelo base. Para operación comunitaria, el sistema transforma esta probabilidad en un índice de prioridad I_i definido por (2):

$$I_i = 100 \times [0.45p_i + 0.25A_i + 0.20T_i + 0.10C_i]. \qquad (2)$$

Los pesos de (2) son supuestos de diseño provisionales, definidos por los autores para la simulación; no proceden de un proceso formal de elicitación de expertos ni de estimación empírica. Se asignó mayor peso a la probabilidad de deterioro (0.45) por representar riesgo futuro estimado; la severidad antropométrica actual (0.25) por su relevancia clínica inmediata; la tendencia longitudinal (0.20) por la relevancia del crecimiento insuficiente como proceso acumulativo [19] y la vulnerabilidad operativa (0.10) como factor de equidad en el acceso a servicios. Antes de cualquier uso operativo, estos valores deberán sustituirse por pesos obtenidos mediante un procedimiento documentado de consenso, por ejemplo un Delphi modificado con registro de desacuerdos, y someterse a un análisis de sensibilidad que cuantifique cuántas asignaciones de prioridad cambian bajo combinaciones alternativas plausibles.

En la Fórmula (2), A_i es severidad antropométrica normalizada (derivada de WHZ y MUAC en t_0), T_i es tendencia longitudinal (pendiente de WHZ en las visitas previas) y C_i es vulnerabilidad operativa (distancia, disponibilidad alimentaria y posibilidad de visita), todas escaladas a [0, 1]. WHZ, MUAC y la pendiente de WHZ también son predictores de p_i, de modo que A_i y T_i reutilizan información ya contenida en la probabilidad. La superposición es deliberada: p_i expresa riesgo futuro, mientras que A_i y T_i mantienen visibles la gravedad y la tendencia actuales, aunque el modelo las pondere poco. Aun así, supone una doble ponderación que el análisis de sensibilidad de los pesos deberá cuantificar. C_i no comparte variables con p_i. La fórmula no reemplaza la regla clínica, porque cualquier alarma roja activa

referencia, aunque el valor de p_i sea bajo. Esta decisión preserva seguridad y reduce el riesgo de omisión por error del modelo.

La Fórmula (1) corresponde al modelo base de regresión logística, que sirve como línea de referencia interpretable. Cuando se emplea XGBoost, la probabilidad p_i se obtiene mediante la transformación sigmoide aplicada a la suma de árboles de decisión potenciados, conservando la misma interfaz de salida probabilística, pero con una función de estimación de mayor capacidad. La Fórmula (2) opera de forma idéntica independientemente del modelo subyacente

El flujo de clasificación se diseñó como proceso con compuertas de decisión. Primero se verifica plausibilidad del dato. Si el registro es implausible, se repite la medición durante la visita. Luego se calculan puntuaciones *z* y perímetro braquial medio. Si hay alarma clínica, se genera referencia inmediata. Si no hay alarma, el modelo estima riesgo y el sistema asigna prioridad operativa. La Fig. 2 resume este flujo. La Tabla II presenta las reglas de salida, su fuente normativa y los tiempos máximos sugeridos de respuesta.

TABLA II
REGLAS OPERATIVAS DE CLASIFICACIÓN NUTRICIONAL Y FUENTE DE CADA CRITERIO

Nivel	Condición resumida	Salida	Tiempo	Fuente
Rojo	Edema, MUAC < 11.5 cm, WHZ < -3 o peligro	Referencia inmediata	Mismo día	C [6]
Naranja	WHZ entre -3 y -2, MUAC 11.5 a 12.5 o riesgo >= 0.70	Visita prioritaria	72 horas	C [6]; D (riesgo y plazo)
Amarillo	Tendencia descendente o riesgo 0.40 a 0.69	Seguimiento intensificado	7 a 14 días	D
Verde	Sin alarma y riesgo < 0.40	Control rutinario	30 días	D
Gris	Dato incompleto o implausible	Repetir medición	Visita actual	C [7], [8]

Nota: C = criterio clínico normativo; D = decisión operativa de diseño propuesta por los autores. Los plazos de los niveles naranja, amarillo y verde son propuestas locales de flujo de trabajo, no umbrales de guía; la referencia el mismo día de los casos rojos se alinea con [6].

La evaluación técnica se planteó con datos sintéticos, no con registros reales de niños. Se generó un escenario simulado de 1,200 niños sintéticos, cada uno con una trayectoria de visitas, para verificar consistencia del flujo, comportamiento de métricas y funcionamiento del tablero. Los valores fueron parametrizados dentro de rangos antropométricos plausibles y se introdujeron escenarios de error de captura, visitas perdidas y deterioro longitudinal.

Los registros fueron generados mediante muestreo paramétrico. La edad se distribuyó uniformemente entre 6 y 59 meses y el sexo con probabilidad 0.50. Las mediciones de peso y talla se muestrearon a partir de las medianas y desviaciones estándar de los estándares OMS por edad y sexo, con perturbación gaussiana ($\sigma = 0.15$ del desvío estándar poblacional) para simular variabilidad comunitaria. Se introdujeron intencionalmente: registros implausibles (8%), visitas perdidas en los 90 días previos (12%), tendencia descendente de WHZ en tres mediciones consecutivas (15%), y presencia de edema bilateral o signos de peligro (5%). La proporción de emaciación severa (WHZ < −3) fue 4% y de emaciación moderada (−3 ≤ WHZ < −2) 10%, distribución coherente con escenarios rurales pero no representativa de ninguna comunidad específica. La generación se realizó con Python 3.11, NumPy 1.26 y SciPy 1.12.

La perturbación se aplicó por variable y no se modelaron explícitamente la covarianza conjunta entre peso, talla, MUAC y morbilidad ni la autocorrelación longitudinal mediante los parámetros LMS de la OMS. En consecuencia, pueden aparecer combinaciones poco realistas desde el punto de vista fisiológico aunque cada variable sea marginalmente plausible. El escenario sirve para probar la lógica del flujo, no para reproducir la estructura conjunta de una población real. Los registros implausibles se dirigieron a la categoría gris y no ingresaron al modelo. Una versión reproducible deberá generar las puntuaciones z con los parámetros LMS, con dependencia explícita entre WHZ y MUAC y trayectorias autocorrelacionadas; el código de generación y sus parámetros se pondrán a disposición como material suplementario.

El objetivo de esta simulación fue probar lógica del sistema, no estimar prevalencia ni desempeño clínico definitivo. La validación real requerirá aprobación ética, consentimiento cuando corresponda, control de calidad antropométrico y evaluación prospectiva conforme a TRIPOD+AI.

Control de fuga de información y circularidad. El resultado Y_i se deriva exclusivamente de mediciones posteriores a t_0, y ninguna variable del seguimiento figura entre los predictores (Tabla I). Sin embargo, el mismo generador define las características basales inyectadas (emaciación, tendencia descendente de WHZ, visitas perdidas) y la propensión al deterioro posterior, de manera que los modelos pueden aprender los supuestos de la simulación en lugar de un evento futuro genuinamente incierto. Esta circularidad estructural infla el desempeño aparente y vuelve parcialmente tautológicas las diferencias entre regresión logística, bosque aleatorio, XGBoost y el flujo híbrido.

El conjunto se dividió a nivel de niño en 70% para entrenamiento (n = 840 niños) y 30% para evaluación (n = 360 niños), con estratificación por el resultado Y_i. Como cada niño aporta una sola visita índice, ningún niño contribuye información a ambos subconjuntos y no existe contaminación por mediciones repetidas; la validación temporal, con entrenamiento en periodos sintéticos anteriores y evaluación

en posteriores, queda pendiente. La selección de hiperparámetros de XGBoost (profundidad máxima, tasa de aprendizaje, número de estimadores y proporción de submuestreo) se realizó mediante búsqueda aleatoria con 100 iteraciones y validación cruzada estratificada de cinco pliegues sobre el subconjunto de entrenamiento, optimizando Brier score como función objetivo [15].

La calibración se realizó mediante escalamiento de Platt, ajustado con predicciones fuera de pliegue obtenidas dentro del subconjunto de entrenamiento, sin utilizar el conjunto de evaluación. Sensibilidad y especificidad se calcularon con un umbral preespecificado de $p \geq 0.40$, que coincide con el límite entre control rutinario y seguimiento intensificado de la Tabla II. En el flujo híbrido, un caso se consideró positivo si cumplía un criterio rojo o naranja de la capa de reglas o si $p \geq 0.40$. Ningún umbral se optimizó sobre el conjunto de evaluación. la Tabla III reúne los elementos de reproducibilidad. Los resultados de la Tabla IV corresponden al conjunto de evaluación retenido.

TABLA III
ELEMENTOS DE REPRODUCIBILIDAD DEL EXPERIMENTO SINTÉTICO

Elemento	Especificación
Unidad de análisis	Niño (n = 1,200), con una visita índice t_0 y al menos dos mediciones previas
Resultado	Y_i: deterioro en (t_0, t_0 + 60 d]; evaluación con el criterio compuesto de caso prioritario
Partición	70/30 por niño (840/360), estratificada por Y_i; sin niños compartidos
Preprocesamiento	Puntuaciones z OMS; estandarización para regresión logística; registros grises fuera del modelo
Desbalance de clases	Sin remuestreo; estratificación en la partición y en la validación cruzada
Ajuste de hiperparámetros	Búsqueda aleatoria (100 iteraciones), validación cruzada estratificada de 5 pliegues, objetivo Brier score
Calibración	Escalamiento de Platt con predicciones fuera de pliegue del entrenamiento
Umbral binario	$p \geq 0.40$, preespecificado; híbrido: criterio rojo/naranja o $p \geq 0.40$
Software	Python 3.11, NumPy 1.26, SciPy 1.12
Material suplementario	Código de generación, rangos y valores seleccionados de hiperparámetros, semillas y versiones de bibliotecas

El sistema fue diseñado con identificación exclusivamente codificada y separación entre datos nominales (bajo custodia local) y datos analíticos (anonimizados). No obstante, la implementación real deberá cumplir con la normativa nacional de protección de datos personales aplicable, incluyendo disposiciones específicas para datos de menores de edad. A nivel internacional, el marco ético de la OMS sobre IA en salud [11] y los principios de la Convención sobre los Derechos del Niño obligan a priorizar el interés superior del menor en cualquier uso de sus datos para fines de vigilancia.

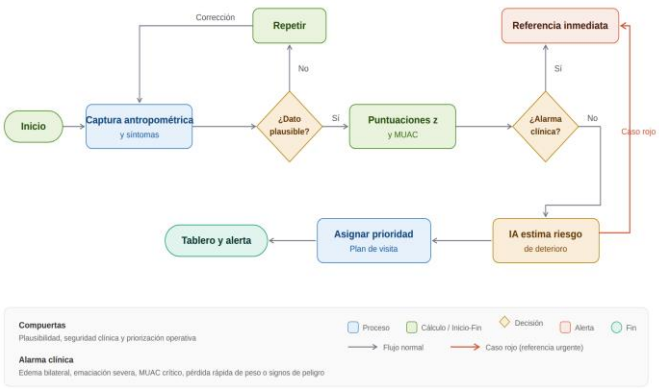


Fig. 2 Flujo estandarizado de clasificación y respuesta comunitaria. El proceso inicia con verificación de plausibilidad, continúa con cálculo de puntuaciones z y evaluación de alarmas clínicas, y concluye con asignación de prioridad operativa (rojo, naranja, amarillo, verde o gris) y acción sugerida con tiempo máximo de respuesta.

IV. RESULTADOS DEL PROTOTIPO

El prototipo conceptual produce cinco salidas: clase de prioridad, explicación breve, acción sugerida, fecha máxima de seguimiento y señal territorial agregada. Las clases son rojo, naranja, amarillo, verde y gris. Rojo corresponde a referencia inmediata. Naranja indica visita prioritaria y plan de seguimiento estrecho. Amarillo señala control intensificado. Verde se asigna a control rutinario. Gris se reserva para mediciones incompletas o implausibles. Esta separación evita que los registros de mala calidad sean tratados como casos normales.

En la simulación técnica, el uso combinado de reglas de seguridad y XGBoost calibrado produjo, frente al criterio compuesto de caso prioritario, una sensibilidad simulada de 0.910 y una especificidad de 0.720, con AUC ROC de 0.860 y Brier score de 0.106. El modelo XGBoost sin reglas obtuvo AUC ROC de 0.870 y Brier score de 0.103, pero menor sensibilidad para casos que debían ser priorizados por seguridad. La regresión logística calibrada mostró desempeño menor, aunque ofreció interpretabilidad directa y utilidad como línea base. El patrón confirma que el flujo se comporta según lo diseñado: la capa de reglas recupera, por construcción, los casos con gravedad ya presente en t_0 que el modelo puede omitir. Como el generador y las reglas comparten supuestos, estas cifras son resultados de prueba de implementación bajo los escenarios preespecificados y no estimaciones del desempeño predictivo clínico. El Brier score, por sí solo, no demuestra calibración [16]; la curva de calibración, el intercepto y la pendiente se reportarán en la réplica con múltiples semillas. La Tabla IV resume los resultados simulados.

TABLA IV
RESULTADOS DE PRUEBA DE IMPLEMENTACIÓN EN EL ESCENARIO SINTÉTICO

Modelo	AUC ROC	Brier	Sensibilidad	Especificidad
Reglas OMS	0.710	0.162	0.760	0.610
Regresión logística	0.780	0.128	0.790	0.690
Bosque aleatorio	0.840	0.115	0.810	0.740
XGBoost calibrado	0.870	0.103	0.820	0.770
XGBoost + reglas	0.860	0.106	0.910	0.720

Nota: Los intervalos de confianza al 95% se estimaron mediante bootstrap con 2,000 remuestreos sobre el conjunto de evaluación. Para el modelo híbrido (XGBoost + reglas): AUC ROC 0.860 (IC 95%: 0.829–0.891), sensibilidad 0.910 (IC 95%: 0.876–0.944), especificidad 0.720 (IC 95%: 0.679–0.761), Brier score 0.106 (IC 95%: 0.091–0.121). Estos intervalos reflejan variabilidad de muestreo dentro de datos sintéticos y no deben interpretarse como estimaciones de desempeño clínico real. Tampoco capturan la incertidumbre estructural del generador; una estimación más informativa exigiría repetir la generación y el entrenamiento con múltiples semillas, tarea reservada para la versión reproducible del experimento. Umbral binario: $p \geq 0.40$ (Tabla III).

El tablero comunitario propuesto consolida indicadores de acción, no solo indicadores descriptivos. Los componentes mínimos son número de niños activos, alertas rojas, visitas vencidas, calidad de dato, distribución por prioridad, priorización territorial y lista de trabajo. La Fig. 3 muestra una maqueta operativa basada en datos sintéticos. El tablero se diseñó para responder tres preguntas de gestión: quién requiere intervención, dónde se concentran los riesgos y qué casos están vencidos. Este enfoque permite que la herramienta sea útil para reuniones comunitarias, planificación semanal de promotores y supervisión municipal. La maqueta evidencia plausibilidad arquitectónica y de diseño de interfaz; la factibilidad operativa y la usabilidad no han sido evaluadas.

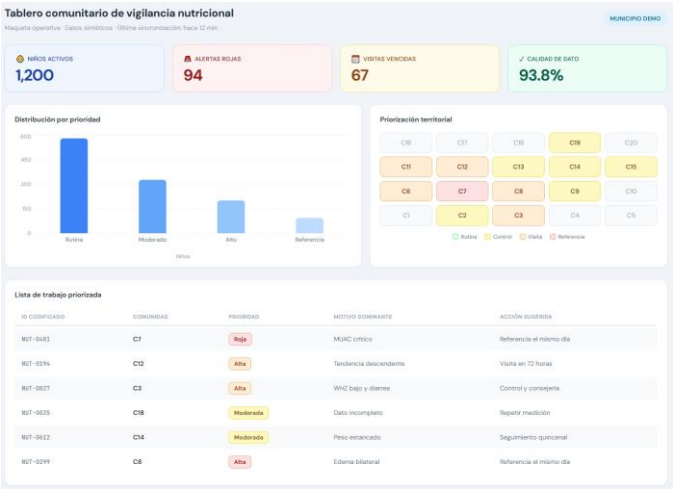


Fig. 3 Maqueta del tablero comunitario de vigilancia nutricional con datos sintéticos. Incluye indicadores de niños activos, alertas por nivel de prioridad, visitas vencidas, calidad de captura y distribución territorial de riesgos.

La trazabilidad del sistema registra versión de reglas, versión del modelo, usuario que captura, usuario que confirma, fecha de alerta y cierre de caso. Cada predicción debe incluir una explicación local, por ejemplo, perímetro braquial bajo, disminución reciente de peso, visita perdida o diarrea reciente. La explicación no debe expresarse como certeza diagnóstica, sino como razón de priorización. En términos de gobernanza, la Tabla V resume riesgos críticos y controles propuestos. La aprobación de nuevas versiones del modelo o de las reglas corresponde a un comité técnico designado por la autoridad sanitaria local, que conserva la versión previa validada para restaurarla de inmediato si la nueva incumple los umbrales preespecificados de calibración o de seguridad.

TABLA V
RIESGOS DE IMPLEMENTACIÓN Y CONTROLES DE GOBERNANZA

Riesgo	Control	Indicador
Error de medición	Doble medición y plausibilidad	% registros implausibles
Sesgo territorial	Monitoreo por comunidad	Diferencia de sensibilidad
Falsa seguridad por IA	Reglas no anulables	Alertas cerradas sin visita
Privacidad	ID codificado y roles	Accesos no autorizados
Deriva del modelo	Revisión programada y recalibración ante umbrales de deriva preespecificados	Cambio en Brier score, intercepto y pendiente

V. DISCUSIÓN

Trabajos previos han explorado aprendizaje automático para predecir malnutrición infantil. Un metaanálisis reciente identificó 11 estudios que aplicaron aprendizaje automático a encuestas demográficas y de salud (DHS), con exactitudes agrupadas de 68.9% para retraso del crecimiento y 84.4% para emaciación [20]. Rahman et al. analizaron los factores de riesgo de retraso del crecimiento, emaciación y bajo peso en Bangladesh y los predijeron con algoritmos de aprendizaje automático [21]. Fenta et al. utilizaron datos de la DHS de Etiopía con múltiples clasificadores para predecir desnutrición por zonas administrativas; el bosque aleatorio fue el mejor modelo, y la residencia urbano-rural y la alfabetización parental resultaron los predictores más relevantes [22]. Estos trabajos operan sobre datos retrospectivos de encuestas poblacionales y no reportan captura en tiempo real, explicabilidad por caso individual ni generación de listas de trabajo para promotores comunitarios. El sistema aquí propuesto se diferencia al vincular predicción con acción operativa dentro de un flujo unificado de vigilancia.

Las herramientas móviles de vigilancia nutricional, por su parte, han privilegiado la captura y la clasificación. Chanani et al. incorporaron a CommCare una calculadora de grado nutricional basada en las tablas OMS y redujeron los errores de clasificación de trabajadores de primera línea [23]; Irache et al. desarrollaron un sistema móvil de recolección para el monitoreo del crecimiento en atención primaria [24]; Jensen et al. evaluaron una aplicación para obtener datos de salud y nutrición de alta frecuencia en poblaciones de difícil acceso

[25], y Bhavnani y Link mostraron que la autocaptura familiar de MUAC en teléfonos inteligentes, con sincronización diferida, alcanza una exactitud de clasificación similar a la de las encuestas en papel [26]. Según lo publicado, ninguno de estos sistemas combina reglas de seguridad no anulables, predicción de riesgo, explicación por caso y listas de trabajo. La Tabla VI sintetiza la comparación. Dado que no procede de una revisión sistemática, la afirmación de novedad se limita a la combinación de funciones y no a cada componente.

TABLA VI
COMPARACIÓN FUNCIONAL CON SISTEMAS Y ESTUDIOS RELACIONADOS

Función	[23]	[24]	[25]	[26]	[20]–[22]	Este trabajo
Captura móvil offline	NR	NR	NR	S	N	S
Vigilancia longitudinal	S	S	S	S	N	S
Reglas de seguridad	P	NR	NR	P	N	S
Predicción de riesgo con ML	N	N	N	N	S	S
Explicación por caso	N	N	N	N	NR	S
Bitácora de auditoría	NR	NR	NR	NR	N	S
Listas de trabajo	NR	NR	NR	NR	N	S
Tablero territorial	NR	NR	NR	NR	P	S
Despliegue en campo	S	S	S	S	N	N

Nota: S = sí; P = parcial; N = no; NR = no reportado en la fuente. Evaluación de los autores a partir de lo publicado. Para este trabajo, S indica funcionalidad diseñada, no validada en campo.

El sistema propuesto aborda una brecha frecuente entre medición nutricional y acción territorial. La innovación no consiste solo en calcular puntuaciones *z*, sino en convertir mediciones comunitarias en un flujo continuo de clasificación, alerta y seguimiento. Conviene precisar que la novedad no es algorítmica: XGBoost, SHAP, el modelo C4, las reglas antropométricas OMS y los tableros son enfoques consolidados; lo nuevo es su orquestación en un flujo de acción comunitaria. En comunidades rurales, donde el costo de desplazamiento es alto y el personal es limitado, una lista de trabajo priorizada podría mejorar el uso de recursos sin depender de infraestructura compleja, hipótesis que deberá contrastarse en campo.

Desde una perspectiva tecnológica, la arquitectura offline primero es relevante para zonas con conectividad intermitente. La separación entre aplicación móvil, motor antropométrico, clasificador, auditoría y tablero permite mantener modularidad y actualizar componentes sin rediseñar todo el sistema. También facilita una ruta de madurez: iniciar con reglas OMS y tablero, agregar modelo logístico calibrado cuando existan datos longitudinales suficientes, y evaluar modelos más complejos únicamente cuando el beneficio sea medible y auditado.

Desde una perspectiva de negocio público, el sistema podría implementarse como plataforma municipal o regional con costos marginales presumiblemente bajos, aún no estimados mediante un modelo formal de costos, una vez que existan dispositivos móviles y capacitación. Su valor se expresa en cobertura de vigilancia, reducción de visitas no priorizadas, detección temprana, trazabilidad de referencias y reportes para gestión local. El retorno esperado no debe medirse solo en ahorro financiero, sino en oportunidad de intervención, continuidad del seguimiento y capacidad de demostrar desempeño del programa ante financiadores o autoridades.

La incorporación de inteligencia artificial exige controles explícitos. En primer lugar, las reglas de seguridad no deben ser anuladas por el modelo. En segundo lugar, la calibración debe evaluarse periódicamente mediante curva de calibración, intercepto y pendiente, y no solo con el Brier score, porque un modelo con alta discriminación puede producir probabilidades mal calibradas. En tercer lugar, la equidad debe analizarse por sexo, edad, comunidad y acceso geográfico, evitando que la herramienta reproduzca desigualdades de captura o seguimiento. En cuarto lugar, el tablero debe mostrar calidad de datos, porque una comunidad con pocos casos puede reflejar baja cobertura y no menor riesgo.

La literatura sobre salud digital documenta una brecha persistente entre el diseño de sistemas y su adopción sostenida en contextos de recursos limitados. Labrique et al. señalaron que la rápida proliferación de proyectos de mHealth ha consistido principalmente en esfuerzos piloto, y propusieron un marco de aplicaciones orientado al fortalecimiento de los sistemas de salud [27]. El sistema propuesto busca mitigar parcialmente este riesgo mediante diseño modular, operación offline y bajo costo marginal, pero su sostenibilidad dependerá de la apropiación institucional por parte de las secretarías de salud y de la integración con los sistemas de información existentes a nivel municipal y regional

El prototipo tiene limitaciones. La validación presentada es simulada y solo demuestra consistencia técnica del diseño. No permite afirmar eficacia clínica, impacto sobre recuperación nutricional ni reducción de mortalidad. Tampoco sustituye la valoración médica, el diagnóstico nutricional completo ni la disponibilidad de servicios de tratamiento. Tampoco se efectuaron pruebas de factibilidad técnica: tasa de éxito de captura y sincronización offline, resolución de conflictos, pérdida de datos, especificación mínima de hardware, tamaño promedio de la carga de sincronización, control de acceso por roles, completitud de la bitácora de auditoría y matriz preliminar de costos. El trabajo demuestra, por tanto, plausibilidad arquitectónica y no factibilidad operativa; esas pruebas constituyen la siguiente etapa preclínica. Además, el generador sintético no modela la dependencia conjunta de las variables y comparte supuestos con la capa de reglas, y los pesos del índice de prioridad son supuestos de diseño. La implementación real requiere protocolos locales, capacitación antropométrica, gestión de consentimiento, seguridad de datos, pruebas de campo y evaluación prospectiva. Además, el modelo debe actualizarse con datos locales auditados antes de

apoyar decisiones programáticas a gran escala.

Una futura validación debe incluir al menos cuatro fases. La primera fase debe comprobar la factibilidad técnica descrita en las limitaciones, la exactitud de captura y concordancia antropométrica entre observadores. La segunda debe evaluar desempeño predictivo con registros longitudinales y validación temporal, partición por niño y calibración reportada conforme a TRIPOD+AI [12]. La tercera debe medir utilidad operativa mediante indicadores de visitas realizadas, tiempos de referencia y casos vencidos. La cuarta debe evaluar impacto comunitario y aceptabilidad por promotores, familias y personal de salud. Solo después de estas fases sería apropiado afirmar que el sistema mejora la vigilancia nutricional continua.

VI. Conclusiones

Se propuso un sistema de estratificación nutricional infantil con inteligencia artificial para comunidades rurales, compuesto por una capa de reglas clínicas, un modelo interpretable de riesgo futuro y un tablero comunitario de vigilancia continua. El diseño integra estándares OMS, captura offline, control de calidad, explicabilidad y trazabilidad. En los escenarios sintéticos de prueba, el flujo híbrido aumentó la sensibilidad simulada para los casos prioritarios predefinidos sin eliminar la supervisión humana; si ello se traduce en una mejor vigilancia en condiciones reales deberá establecerse mediante validación prospectiva. El principal aporte del trabajo es un modelo de integración y orquestación operativa para transformar mediciones comunitarias en acciones priorizadas. Su valor definitivo deberá confirmarse mediante validación prospectiva, análisis de equidad, evaluación de utilidad y gobernanza de datos en condiciones reales de atención rural.

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] UNICEF, World Health Organization, and World Bank Group, *Levels and Trends in Child Malnutrition: UNICEF/WHO/World Bank Group Joint Child Malnutrition Estimates, Key Findings of the 2025 Edition*. New York, NY, USA: UNICEF, 2025. [Online]. Available: https://data.unicef.org/resources/jme/. Accessed: May 14, 2026.

[2] World Health Organization, "Joint child malnutrition estimates (UNICEF-WHO-WB)", *WHO Global Health Observatory*, 2025. [Online]. Available: https://www.who.int/data/gho/data/themes/topics/joint-child-malnutrition-estimates-unicef-who-wb. Accessed: May 14, 2026.

[3] Development Initiatives, "Honduras: Country nutrition profile", *Global Nutrition Report*, 2024. [Online]. Available: https://globalnutritionreport.org/resources/nutrition-profiles/latin-america-and-caribbean/central-america/honduras/. Accessed: May 14, 2026.

[4] World Health Organization, *WHO Child Growth Standards: Length/Height-for-Age, Weight-for-Age, Weight-for-Length, Weight-for-Height and Body Mass Index-for-Age: Methods and Development*. Geneva, Switzerland: WHO, 2006. [Online]. Available: https://www.who.int/publications/i/item/924154693X. Accessed: May 14, 2026.

[5] WHO Multicentre Growth Reference Study Group, "WHO Child Growth Standards based on length/height, weight and age", *Acta Paediatr. Suppl.*, vol. 450, pp. 76–85, Apr. 2006, doi: 10.1111/j.1651-2227.2006.tb02378.x.

[6] World Health Organization, *WHO Guideline on the Prevention and Management of Wasting and Nutritional Oedema (Acute Malnutrition) in Infants and Children Under 5 Years*. Geneva, Switzerland: WHO, 2023. [Online]. Available: https://www.who.int/publications/i/item/9789240082830. Accessed: May 14, 2026.

[7] World Health Organization, "WHO Anthro Survey Analyser and other tools", *WHO Child Growth Standards Software*, 2019. [Online]. Available: https://www.who.int/tools/child-growth-standards/software. Accessed: May 14, 2026.

[8] SMART Initiative, *SMART Methodology Manual 2.0*. Geneva, Switzerland: SMART Initiative, 2017. [Online]. Available: https://smartmethodology.org/survey-planning-tools/smart-methodology/smart-methodology-manual/. Accessed: May 14, 2026.

[9] World Health Organization, *WHO Guideline: Recommendations on Digital Interventions for Health System Strengthening*. Geneva, Switzerland: WHO, 2019. [Online]. Available: https://www.who.int/publications/i/item/9789241550505. Accessed: May 14, 2026.

[10] World Health Organization, *Classification of Digital Health Interventions v1.0: A Shared Language to Describe the Uses of Digital Technology for Health*. Geneva, Switzerland: WHO, 2018. [Online]. Available: https://www.who.int/publications/i/item/WHO-RHR-18.06. Accessed: May 14, 2026.

[11] World Health Organization, *Ethics and Governance of Artificial Intelligence for Health: WHO Guidance*. Geneva, Switzerland: WHO, 2021. [Online]. Available: https://www.who.int/publications/i/item/9789240029200. Accessed: May 14, 2026.

[12] 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.

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

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

[15] G. W. Brier, "Verification of forecasts expressed in terms of probability", *Mon. Weather Rev.*, vol. 78, no. 1, pp. 1–3, Jan. 1950, doi: 10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2.

[16] B. Van Calster, D. J. McLernon, M. van Smeden, L. Wynants, and E. W. Steyerberg, "Calibration: The Achilles heel of predictive analytics", *BMC Med.*, vol. 17, Art. no. 230, 2019, doi: 10.1186/s12916-019-1466-7.

[17] 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.

[18] S. Brown, "The C4 model for visualising software architecture", *C4 Model*, 2024. [Online]. Available: https://c4model.com/. Accessed: May 14, 2026.

[19] C. P. Stewart, L. Iannotti, K. G. Dewey, K. F. Michaelsen, and A. W. Onyango, "Contextualising complementary feeding in a broader framework for stunting prevention", *Matern. Child Nutr.*, vol. 9, no. S2, pp. 27–45, Sep. 2013, doi: 10.1111/mcn.12088.

[20]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.

[21]S. M. J. Rahman et al., "Investigate the risk factors of stunting, wasting, and underweight among under-five Bangladeshi children and its prediction based on machine learning approach", *PLoS ONE*, vol. 16, no. 6, Art. no. e0253172, 2021, doi: 10.1371/journal.pone.0253172.

[22]H. M. Fenta, T. Zewotir, and E. K. Muluneh, "A machine learning classifier approach for identifying the determinants of under-five child undernutrition in Ethiopian administrative zones", *BMC Med. Inform. Decis. Mak.*, vol. 21, Art. no. 291, 2021, doi: 10.1186/s12911-021-01652-1.

[23]S. Chanani, J. Wacksman, D. Deshmukh, S. Pantvaidya, A. Fernandez, and A. Jayaraman, "M-Health for improving screening accuracy of acute malnutrition in a community-based management of acute malnutrition program in Mumbai informal settlements", *Food Nutr. Bull.*, vol. 37, no. 4, pp. 504–516, 2016, doi: 10.1177/0379572116657241.

[24]A. Irache, R. Murachpersad, and R. Caleyachetty, "The development and application of a mobile-based data collection system for a growth monitoring programme in selected primary care centres in the Republic of Mauritius", *BMJ Glob. Health*, vol. 4, no. 6, Art. no. e001928, 2019, doi: 10.1136/bmjgh-2019-001928.

[25]N. Jensen, W. Lepariyo, V. Alulu, S. Sibanda, and B. Kiage, "Assessing Mbiotisho: A smartphone application used to collect high-frequency health and nutrition data from difficult-to-reach populations", *Matern. Child Nutr.*, vol. 19, Art. no. e13496, 2023, doi: 10.1111/mcn.13496.

[26]R. Bhavnani and N. S. Link, "D2A: A community-led smartphone tool for malnutrition screening in Kenya", *Front. Public Health*, vol. 13, Art. no. 1695850, 2025, doi: 10.3389/fpubh.2025.1695850.

[27]A. B. Labrique, L. Vasudevan, E. Kochi, R. Fabricant, and G. Mehl, "mHealth innovations as health system strengthening tools: 12 common applications and a visual framework," *Global Health: Sci. Pract.*, vol. 1, no. 2, pp. 160–171, Aug. 2013, doi: 10.9745/GHSP-D-13-00031.