Regional Digital Surveillance Platform for Arboviruses Using Machine Learning and Community Data

Yolly Molina[1]; Patricia Soriano[2]; Saulo Romero[3]
[1,3]Universidad Nacional Autónoma de Honduras, Honduras, *yolly.molina@unah.edu.hn, saulo.romero@unah.edu.hn*
[2]Ministerio Público de Honduras, Honduras, *paty.soriano@yahoo.com*

Abstract– This study aimed to demonstrate the feasibility of a regional digital surveillance platform for arboviruses in Latin America through secondary analysis of OpenDengue V1.3 and a machine learning prototype, using dengue as a tracer disease. A total of 29,873 national records from 129 countries were filtered to 11,534 records from 29 Latin American and Caribbean countries, and 2010-2024 indicators were calculated. Persistence, Ridge, Random Forest and Gradient Boosting models were trained with temporal split 1995-2020 and test 2021-2024. The region accumulated 41.8 million cases between 2010-2024, with an all-time high of 13.0 million in 2024. Honduras reported 177,209 cases in 2024 (1,688 per 100,000 inhabitants). The binary outbreak classifier reached AUC 0.702. OpenDengue's open data allows building a reproducible regional architecture that integrates official data, citizen science, and machine learning to strengthen the response to the largest documented epidemic in the region. A three-phase pilot validation protocol and modelling improvement strategies based on climatic, entomological and mobility covariates are also proposed.

Keywords-- arboviruses, epidemiological surveillance, machine learning, open data, Latin America.

.

Plataforma regional de vigilancia digital para arbovirosis usando aprendizaje automático y datos comunitarios

Yolly Molina[1]; Patricia Soriano[2]; Saulo Romero[3]
[1,3]Universidad Nacional Autónoma de Honduras, Honduras, *yolly.molina@unah.edu.hn, saulo.romero@unah.edu.hn*
[2]Ministerio Público de Honduras, Honduras, *paty.soriano@yahoo.com*

Resumen– ***El estudio tuvo como objetivo demostrar la viabilidad de una plataforma regional de vigilancia digital para arbovirosis en América Latina mediante análisis secundario de OpenDengue V1.3 y un prototipo de aprendizaje automático, con el dengue como enfermedad trazadora. Se tomaron 29,873 registros nacionales de 129 países, que se filtraron a 11,534 registros de 29 países latinoamericanos y caribeños, y se calcularon indicadores 2010-2024. Se entrenaron modelos de persistencia, Ridge, Random Forest y Gradient Boosting con split temporal 1995-2020 y prueba 2021-2024. La región acumuló 41.8 millones de casos entre 2010-2024, con máximo histórico de 13.0 millones en 2024. Honduras reportó 177,209 casos en 2024 (1,688 por 100,000 habitantes). El clasificador binario de brotes alcanzó AUC 0.702. Los datos abiertos de OpenDengue permiten construir una arquitectura regional reproducible que integra datos oficiales, ciencia ciudadana y aprendizaje automático para fortalecer la respuesta ante la mayor epidemia documentada en la región. Se proponen además un protocolo de validación piloto en tres fases y estrategias de mejora del modelado basadas en covariables climáticas, entomológicas y de movilidad.***

Palabras clave-- arbovirosis, vigilancia epidemiológica, aprendizaje automático, datos abiertos, América Latina.

I. Introducción

El dengue es la arbovirosis humana de mayor expansión global y representa una carga creciente para los sistemas sanitarios de América Latina y el Caribe. La estimación de Bhatt y colaboradores ya señalaba 390 millones de infecciones anuales en el mundo, con la región latinoamericana entre los focos prioritarios de transmisión [1]. La Organización Mundial de la Salud (OMS) advierte que la incidencia mundial se ha multiplicado por más de ocho en las últimas dos décadas, con la mitad de la población mundial en riesgo [2]. La Organización Panamericana de la Salud (OPS) declaró en 2024 una alerta epidemiológica regional ante un aumento sin precedentes de casos, que superó los 11.7 millones para la semana epidemiológica 36 y que terminó el año con cifras superiores a 13 millones, casi tres veces el récord previo de 2023 [3], [4]. La OPS atribuyó este escenario a una combinación de factores que incluye la expansión geográfica de Aedes aegypti, la cocirculación simultánea de los cuatro serotipos en al menos siete países y la influencia del fenómeno El Niño bajo un contexto de cambio climático [5].

La vigilancia epidemiológica tradicional, basada en notificación pasiva y reporte semanal agregado, presenta retrasos estructurales que dificultan la respuesta oportuna durante brotes explosivos. Codeço y colaboradores documentaron desde 2018 los retos de articular sistemas de información para arbovirosis, con la propuesta del nowcasting Infodengue como ejemplo de operación municipal en Brasil [6]. Más recientemente, el mismo grupo describió la rápida expansión territorial del dengue en Brasil, con presión transmisora creciente sobre nuevas áreas geográficas [7]. La iniciativa OpenDengue, publicada en Scientific Data en 2024, agregó a este panorama una infraestructura de datos abiertos comparable y reutilizable, con más de 56 millones de casos estandarizados de 102 países en su versión 1.2 y una cobertura ampliada a 129 países en la versión 1.3 publicada en mayo de 2025 [8], [9].

De manera paralela, la inteligencia artificial aplicada a salud pública ha mostrado utilidad creciente en la predicción de brotes. Sebastianelli y colaboradores desarrollaron un modelo de ensamble reproducible para pronosticar la incidencia de dengue a un mes en Brasil y Perú [10]. Cabrera y colaboradores revisaron la aplicación de aprendizaje automático al dengue en América Latina bajo el enfoque One Health, identificando una concentración de estudios en Brasil, Colombia y México y un uso predominante de variables meteorológicas y climáticas [11]. Una revisión narrativa actualizada publicada en 2026 confirmó que los algoritmos Random Forest y Long Short-Term Memory ofrecen el mejor desempeño en horizontes cortos de hasta una semana, aunque pocos sistemas pasan de la fase académica a la operación rutinaria [12]. Lowe y colaboradores demostraron además que el riesgo de dengue en Brasil aumenta tras eventos extremos hidrometeorológicos, lo que respalda la integración de variables climáticas en sistemas de alerta [13].

La ciencia ciudadana ha emergido como un componente complementario relevante. Plataformas como Mosquito Alert, iNaturalist y GLOBE Observer Mosquito Habitat Mapper han demostrado que es posible recolectar a gran escala observaciones georreferenciadas de criaderos y vectores con teléfonos inteligentes, y entrenar modelos de aprendizaje profundo capaces de identificar especies con desempeño cercano al de los entomólogos [14]. Sin embargo, América

Latina aún carece de una arquitectura regional integrada que articule datos oficiales abiertos, datos comunitarios y modelos predictivos bajo principios FAIR y de soberanía de datos.

El presente trabajo aborda esta brecha mediante un estudio de caso que combina dos contribuciones. Primero, un análisis secundario empírico de OpenDengue V1.3 para América Latina y el Caribe, con foco en la epidemia regional de 2024. Segundo, una propuesta arquitectónica concreta de plataforma regional de vigilancia digital para arbovirosis que integra datos comunitarios, modelado predictivo con aprendizaje automático y mecanismos de gobernanza alineados con el Reglamento Sanitario Internacional. El objetivo general es demostrar que los datos abiertos disponibles permiten construir prototipos reproducibles que pueden orientar decisiones de salud pública en escenarios de transmisión explosiva como el observado en 2024. Aunque la arquitectura se concibe para el conjunto de arbovirosis de relevancia regional, el análisis empírico se centra en el dengue por ser la única que dispone de una base regional abierta, estandarizada y con series largas; el dengue opera así como enfermedad trazadora para validar componentes extensibles a chikungunya, Zika y Oropouche.

II. Materiales y métodos

A. Tipo de estudio y decisión metodológica

El estudio se diseñó como análisis secundario de datos abiertos con un componente complementario de propuesta arquitectónica. El análisis empírico de fuentes abiertas tiene prioridad sobre revisiones documentales cuando existen datos descargables y verificables. La selección de OpenDengue V1.3 como fuente principal respondió a cuatro criterios. Primero, la disponibilidad de archivos CSV completos en el repositorio público GitHub OpenDengue/master-repo bajo licencia Creative Commons CC BY-SA 4.0 [8], [9]. Segundo, la cobertura temporal y geográfica adecuada, con datos hasta abril de 2025 para 121 países. Tercero, la trazabilidad documentada de cada registro mediante UUID al expediente fuente. Cuarto, la cocita oficial PAHO-PLISA y OMS Global Dengue Dashboard que respalda la calidad técnica del extracto.

Se descartaron como fuentes principales otras alternativas. Project Tycho, aunque útil históricamente, no incluye datos posteriores a 2012 para América Latina y solo dos países cuentan con resolución administrativa de segundo nivel [8]. PAHO PLISA es la fuente primaria oficial pero no ofrece descarga estandarizada en archivo único, lo que dificulta análisis transnacionales reproducibles [15]. OMS Global Dengue Dashboard se incorporó como fuente complementaria, no principal, debido a la asimetría de su cobertura por país.

B. Fuente de datos principal

Se utilizó la versión 1.3 de OpenDengue, publicada el 27 de mayo de 2025 [8], [9]. Se descargaron dos archivos: National_extract_V1_3.csv, con 29,873 registros nacionales, y Temporal_extract_V1_3.csv, con 2,835,031 registros nacionales y subnacionales con la mejor resolución temporal disponible. La consulta del repositorio se realizó el 7 de mayo de 2026.

C. Criterios de inclusión y limpieza

Del extracto nacional se filtraron los registros correspondientes a 29 países de América Latina y el Caribe (LAC), conservando 11,534 observaciones válidas con valores no nulos en las variables Year y dengue_total y con conteos no negativos. La unidad de análisis es la observación país-año-fuente. Se aplicó agregación por país y año, lo que produjo una serie temporal nacional comparable. Para los análisis predictivos se conservaron 23 países con al menos seis observaciones anuales consecutivas, condición indispensable para la construcción de las variables retardadas.

D. Variables de analizadas

Las variables descriptivas incluyeron casos anuales reportados por país, casos acumulados 2010-2024, casos acumulados 2020-2024, tasa de incidencia 2024 por 100,000 habitantes (denominador poblacional aproximado del año 2022 obtenido de fuentes oficiales nacionales), variabilidad inter-anual mediante coeficiente de variación, y razón de casos 2024 sobre la media de casos 2019-2023. Para el modelado predictivo se construyeron las siguientes variables explicativas: lag1 a lag5 de casos en escala log(1+x), media de los cinco años previos, máximo de los cinco años previos, tendencia entre lag3 y lag1, razón lag1 sobre la media móvil, y un identificador numérico de país como variable categórica.

E. Procedimiento estadístico y computacional

Todo el análisis se ejecutó en Python 3 con las librerías pandas para manipulación de datos y scikit-learn para los modelos predictivos [16]. Se usó división temporal train 1995-2020 y test 2021-2024 para evitar fuga de información. La variable objetivo principal en regresión fue log(1 + casos anuales). Se compararon cinco modelos: línea base de persistencia (predecir casos del año t como casos en t-1), media móvil de cinco años, regresión lineal Ridge con regularización L2, Random Forest y Gradient Boosting. Para la detección de brotes se construyó una variable binaria que codifica si los casos de un año superan en un 50 por ciento la media de los cinco años previos. La clasificación binaria se entrenó con un Random Forest balanceado con 500 árboles y profundidad máxima de 10, evaluado mediante exactitud, precisión, sensibilidad, F1 y AUC.

E. Consideraciones éticas y reproducibilidad

El estudio utilizó exclusivamente datos agregados públicos sin información identificable. La licencia CC BY-SA 4.0 de OpenDengue permite reutilización con atribución y redistribución bajo la misma licencia [8]. La arquitectura propuesta se diseñó conforme a los principios FAIR, que postulan que los datos científicos deben ser localizables, accesibles, interoperables y reutilizables [17]. El código

completo del análisis y de la generación de figuras se conservó en archivos Python ejecutables, con semillas aleatorias fijadas para garantizar reproducibilidad. El flujo completo se sintetiza en la Fig. 1.

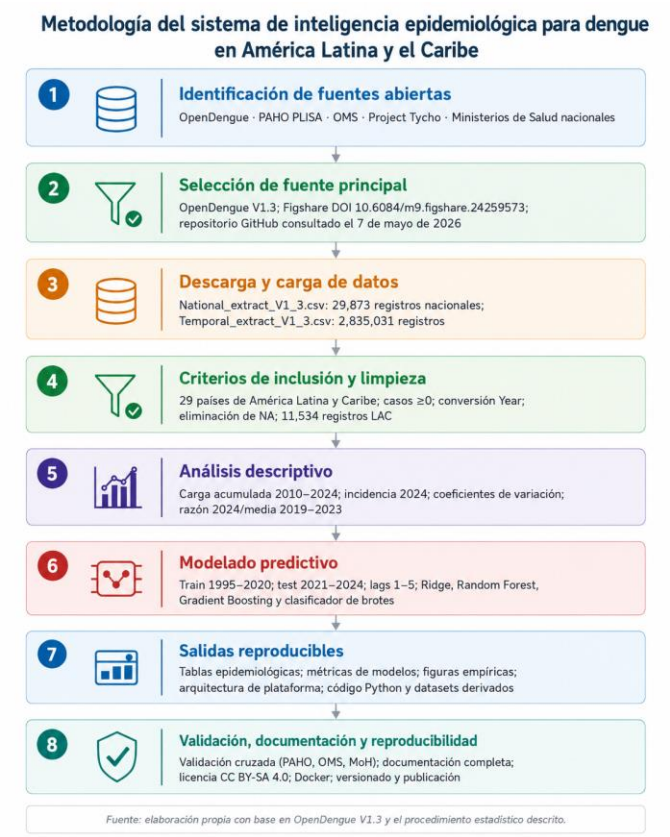


Fig. 1 Flujo metodológico para la identificación, descarga, limpieza y análisis de datos de OpenDengue V1.3 con foco en América Latina y el Caribe.

III. RESULTADOS

A. *Caracterización del dataset analizado*

El extracto nacional de OpenDengue V1.3 contenía 29,873 registros y abarcaba 129 países entre 1924 y 2025. Tras aplicar el filtro de América Latina y el Caribe se conservaron 11,534 registros provenientes de 29 países, con un total acumulado de 53.7 millones de casos reportados a lo largo de toda la serie histórica. La Tabla I presenta las fuentes abiertas consideradas y la justificación de su uso o descarte.

TABLA I
FUENTES DE DATOS ABIERTAS CONSIDERADAS PARA EL ESTUDIO.

Fuente	Tipo de dato	Acceso	Variables relevantes	Estado
OpenDengue V1.3 [8], [9]	Casos agregados nacionales y subnacionales	Sí, CSV abierto	Casos, fechas, año, ISO país, codigos GAUL, definición de caso	Usada (fuente principal)
PAHO PLISA Dengue Indicators [15]	Indicadores semanales y anuales	Parcial, sin descarga única	Casos sospechosos, confirmados, defunciones, serotipos	Usada como fuente complementaria
OMS Global Dengue Dashboard [2]	Indicadores agregados anuales	Sí, panel interactivo	Casos, defunciones, agregados regionales	Usada como referencia
Project Tycho [8]	Casos históricos hasta 2012	Sí, descarga abierta	Casos, año, semana, país	Descartada (no actualizada)
Infodengue Brasil [6]	Nowcasting municipal	Sí, API REST	Casos semanales, alerta por color, Rt	Citada como referencia operativa
Mosquito Alert [14]	Datos comunitarios de vectores	Sí, GBIF dataset	Reportes geolocalizados con foto y validación experta	Considerada para componente comunitario

B. *Carga regional 2010-2024*

La región acumuló 41.8 millones de casos reportados entre 2010 y 2024 según los datos analizados. La serie muestra un patrón cíclico irregular con picos en 2013 (2.4 millones), 2015 (2.5 millones), 2016 (2.9 millones) y 2019 (3.2 millones), seguidos de un valle en 2017-2018 (0.9 y 0.8 millones respectivamente). El año 2024 marcó un punto de inflexión con 13.0 millones de casos analizados, magnitud equivalente a 2.8 veces el récord previo de 2023 (4.7 millones). Estas cifras son consistentes con las alertas oficiales de la OPS, que reportó 11.7 millones para la semana 36 [3] y proyectó cifras superiores al cierre del año, en lo que se ha convertido en la mayor epidemia regional desde que existen registros sistemáticos [4], [18]. El Panel A de la Fig. 2 muestra la tendencia anual y el Panel C contrasta los casos 2024 con la media 2019-2023 para diez países seleccionados.

C. *Distribución por país e incidencia 2024*

Para el año 2024 los países con mayor incidencia por 100,000 habitantes fueron Guyana (5,731), Guayana Francesa (5,292), Brasil (4,769), Paraguay (4,350), Honduras (1,688), Nicaragua (1,334) y Argentina (1,270). Brasil concentró el 78.7 por ciento de los casos absolutos del año (10.27 millones), seguido por Argentina (581,559), México (558,846), Colombia (320,982) y Perú (280,726). La razón entre los casos 2024 y la media 2019-2023 superó el factor de 13 en Argentina, 6.6 en Guatemala, 5 en Brasil, 4.3 en Paraguay y 3.5 en Colombia y 3.0 Perú, lo que confirma una intensificación regional ampliamente sincronizada, aunque no universal: Nicaragua y Bolivia registraron razones inferiores a 1 tras epidemias previas de gran magnitud. La Tabla II sintetiza los principales indicadores por país.

D. *Volatilidad inter-anual*

El coeficiente de variación inter-anual entre 2010 y 2024 superó el 200 por ciento en países como Argentina (249 por ciento), Bahamas (332 por ciento) y Uruguay (266 por ciento), reflejo de su carácter epidémico irregular. En contraste, países con transmisión más sostenida como Brasil (122 por ciento), Paraguay (122 por ciento) y Perú (129 por ciento) mostraron variabilidad alta, pero con base media elevada. Esta heterogeneidad respalda la necesidad de modelos que

ponderen explícitamente la dinámica histórica y la incertidumbre territorial.

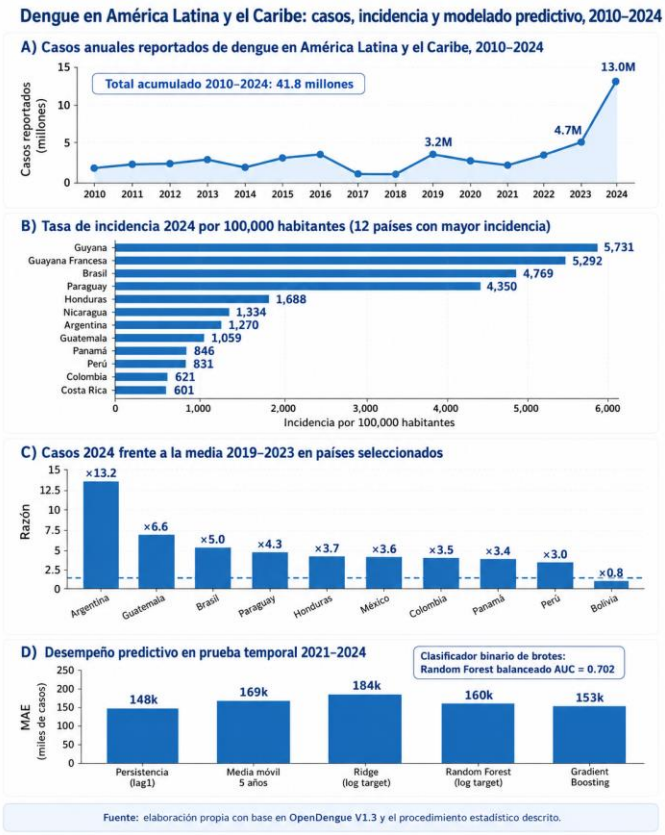


Fig. 2 Resultados empíricos derivados del análisis de OpenDengue V1.3 para América Latina y el Caribe. (A) Casos anuales reportados 2010-2024. (B) Tasa de incidencia 2024 por 100,000 habitantes para los doce países más afectados. (C) Comparación de casos 2024 frente a la media 2019-2023 en escala logarítmica para países seleccionados. (D) Desempeño predictivo del Random Forest y la línea base de persistencia en el conjunto de prueba 2021-2024.

TABLA II
INDICADORES EPIDEMIOLÓGICOS DERIVADOS DEL ANÁLISIS DE OPENDENGUE V1.3 PARA PAÍSES SELECCIONADOS DE AMÉRICA LATINA Y EL CARIBE

País	Casos 2024	Casos acumulados 2020-2024	Incidencia 2024 por 100,000	Razón 2024 / media 2019-2023
Brasil	10,267,077	18,220,900	4,769	5.0
Argentina	581,559	797,755	1,270	13.2
México	558,846	1,057,760	438	3.6
Colombia	320,982	658,323	621	3.5
Perú	280,726	735,126	831	3.0
Paraguay	295,785	623,064	4,350	4.3
Guatemala	188,585	280,147	1,059	6.6
Honduras	177,209	282,240	1,688	3.7
Nicaragua	92,022	462,899	1,334	0.8
Bolivia	50,439	350,453	413	0.8
Costa Rica	31,259	85,539	601	2.0
Panamá	37,231	78,264	846	3.4

Fuente: análisis propio sobre OpenDengue V1.3 [8]. Los denominadores poblacionales corresponden a estimaciones nacionales aproximadas para 2022.

E. Modelado predictivo

Se generaron 997 observaciones país-año entre 1965 y 2024 para 23 países con cobertura suficiente. La división temporal asignó 907 observaciones al conjunto de entrenamiento (1995-2020) y 90 al conjunto de prueba (2021-2024). En la prueba, dominada por la epidemia de dengue de 2024, la línea base de persistencia mostró el mejor desempeño en regresión (MAE = 147,882 casos; $R^2 = 0.541$), seguida por la regresión Ridge en escala logarítmica (MAE = 183,682; $R^2 = 0.263$) y por Random Forest (MAE = 187,486; $R^2 = 0.187$). El error logarítmico medio absoluto, métrica menos sensible a los valores atípicos del año 2024, fue de 1.014 para persistencia, 1.093 para Ridge y 1.099 para Random Forest, lo que indica que los modelos no lineales no superaron a las líneas base en este horizonte explosivo. La Tabla III reporta las métricas comparativas.

TABLA III
MÉTRICAS COMPARATIVAS DE LOS MODELOS PREDICTIVOS EN EL CONJUNTO DE PRUEBA 2021-2024

Modelo	MAE	RMSE	R^2	MAE log(1+x)
Persistencia (lag1)	147,882	775,261	0.541	1.014
Media móvil 5 años	169,082	899,586	0.382	1.102
Regresión Ridge (log target)	183,682	982,366	0.263	1.093
Random Forest (log target)	187,486	1,032,299	0.187	1.099
Gradient Boosting (log target)	189,526	1,026,478	0.196	1.210

Nota: MAE: error absoluto medio. RMSE: raíz del error cuadrático medio. R^2: coeficiente de determinación. MAE log(1+x): error absoluto medio sobre la transformación logarítmica.

En la tarea complementaria de clasificación binaria de brotes, el clasificador Random Forest balanceado alcanzó AUC de 0.702 en el conjunto de prueba, con sensibilidad de 0.366, precisión de 0.682 y F1 de 0.476. La línea base, que predice brote cuando lag1 supera la media móvil de cinco años, obtuvo F1 de 0.651 con mayor sensibilidad, pero menor precisión. La importancia relativa de las variables en el clasificador colocó a la razón lag1 sobre media (13.4 por ciento) y a las tendencias de dos y tres años (11.0 por ciento y 9.7 por ciento) como las señales más informativas, por encima incluso del valor absoluto de los retardos. Este patrón es consistente con la literatura, que destaca el valor predictivo de las dinámicas de cambio sobre los niveles absolutos [10], [12], [13].

F. Arquitectura propuesta de la plataforma

Sobre la base del análisis empírico se propone una arquitectura conceptual organizada en cuatro capas operativas y una capa transversal de gobernanza, sintetizada en la Fig. 3. La capa de fuentes integra OpenDengue, PLISA, OMS, datos comunitarios y variables ambientales. La capa de ingesta normaliza los flujos mediante geomatching con códigos GAUL e ISO 3166, conversión de cumulativos a incidentes y validación cruzada. La capa de almacenamiento y aprendizaje automático aloja el data lake, el feature store y los modelos predictivos. La capa de salidas ofrece un dashboard regional,

alertas tempranas, una API abierta y materiales para usuarios institucionales y comunitarios. La capa transversal de gobernanza articula acuerdos de uso de datos, anonimización, ética algorítmica y sostenibilidad.

La arquitectura es agnóstica respecto de la enfermedad: cada registro se indexa por evento de notificación, lo que permite incorporar chikungunya, Zika y Oropouche como módulos que comparten la capa de ingesta y las covariables ambientales y, en las arbovirosis transmitidas por Aedes, el mismo componente entomológico y comunitario. La reactividad serológica cruzada entre flavivirus y la superposición clínica obligan, no obstante, a que las alertas específicas por arbovirosis se apoyen en confirmación de laboratorio.

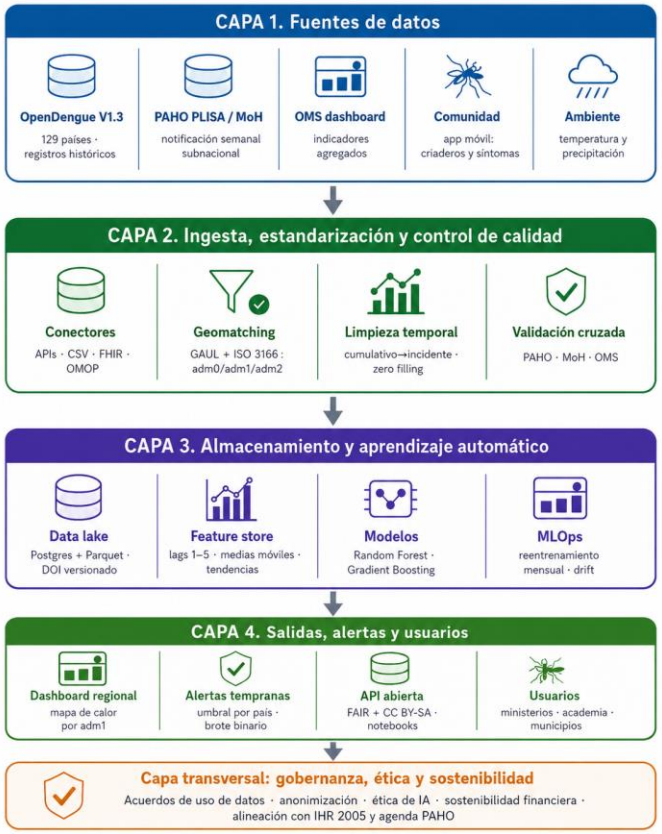


Fig. 3 Arquitectura propuesta de la plataforma regional de vigilancia digital para arbovirosis. La capa transversal de gobernanza, ética y sostenibilidad alinea la propuesta con el Reglamento Sanitario Internacional 2005 y la agenda regional de la OPS.

G. *Protocolo de validación piloto propuesto*

La plataforma aún no se ha probado en un entorno operativo, por lo que su viabilidad se evaluará mediante un piloto escalonado en tres fases con criterios de avance definidos a priori. La fase 1, retrospectiva, reproducirá en modo de pronóstico semanal las temporadas 2019-2024 con datos subnacionales del extracto temporal de OpenDengue y de PLISA para Honduras, Nicaragua y Paraguay, países con incidencias elevadas en 2024 y sistemas de notificación de escala manejable. Se medirán la puntuación de intervalo ponderada (WIS), la cobertura de los intervalos de predicción al 90 por ciento, la sensibilidad y la tasa de falsas alarmas frente al canal endémico, y las semanas de anticipación de cada alerta. Se avanzará solo si el sistema iguala o supera a la persistencia en el WIS en al menos dos tercios de las ventanas evaluadas y anticipa el cruce del umbral epidémico con dos o más semanas.

La fase 2, prospectiva y silenciosa, ejecutará la plataforma durante una temporada completa en paralelo con la vigilancia oficial de dos o tres municipios urbanos por país, sin intervenir en las decisiones, y pondrá a prueba el componente comunitario. Sus indicadores serán los reportes por cada 1,000 habitantes, la proporción validada por técnicos de control vectorial, el tiempo entre reporte y validación, la retención de participantes a tres y seis meses y la concordancia entre focos reportados y conglomerados de casos. La fase 3, operativa, integrará las alertas en los flujos de decisión de las autoridades sanitarias y evaluará el tiempo entre alerta y acción, la aceptabilidad entre usuarios institucionales, el costo por alerta útil y la sostenibilidad técnica. El paso entre fases quedará condicionado al cumplimiento de sus criterios y a la formalización de acuerdos de uso de datos con los ministerios de salud.

IV. Discusión

El presente análisis confirma que 2024 fue el año de mayor transmisión de dengue jamás documentado en América Latina y el Caribe, con 13.0 millones de casos en el extracto nacional de OpenDengue V1.3, magnitud coherente con las alertas regionales de la OPS [3], [4], [18], [19]. La distribución de la carga regional revela tres patrones complementarios. Primero, una concentración de casos absolutos en Brasil, México, Argentina, Colombia y Perú, países con sistemas de vigilancia más maduros y mayor capacidad de notificación. Segundo, picos de incidencia por 100,000 habitantes en territorios pequeños como Guyana, Guayana Francesa, Paraguay y Honduras, donde la presión sobre los sistemas locales puede ser proporcionalmente más severa. Tercero, una intensificación sincronizada manifestada en razones 2024 sobre media 2019-2023 iguales o superiores a 3 en nueve de los doce países de la Tabla II, lo que sugiere un fenómeno regional ligado a determinantes climáticos comunes.

Estos hallazgos refuerzan la evidencia sobre el papel del cambio climático y la urbanización en la transmisión del dengue. Heath y colaboradores proyectaron, mediante modelos de árboles potenciados, aumentos de hasta 32 por ciento en la densidad de Aedes aegypti en Brasil para 2080 bajo escenarios de altas emisiones [20]. Lowe y colaboradores demostraron que las sequías extremas elevan el riesgo de dengue tres a cinco meses después en zonas urbanas, y que las condiciones extremadamente lluviosas tienen efecto inmediato

y hasta tres meses [13]. La cocirculación simultánea de los cuatro serotipos en al menos siete países latinoamericanos, documentada por la OPS, añade un riesgo adicional de dengue grave por inmunidad heteróloga [5].

El desempeño predictivo observado debe interpretarse con cautela. La superioridad relativa de la línea base de persistencia en el conjunto de prueba dominado por 2024 ilustra una limitación general de los modelos basados en historia anual sin variables exógenas. Costa y colaboradores describieron en su revisión narrativa actualizada que los modelos Random Forest y Long Short-Term Memory ofrecen mejor desempeño en horizontes cortos, principalmente cuando se incorporan variables meteorológicas, climáticas y entomológicas [12]. Colón-González y colaboradores mostraron en Vietnam que los superensembles probabilísticos que combinan pronósticos climáticos estacionales con casos retardados superan en calibración a los modelos individuales [21]. El ejercicio Infodengue-Mosqlimate 2024 reportó precisamente que ningún modelo destacó consistentemente durante las condiciones atípicas de 2024 impulsadas por el cambio climático [23]. Esto justifica la propuesta arquitectónica de incorporar variables ambientales y datos comunitarios en la próxima iteración del prototipo.

Superar la línea base en años extremos exige cambios en los datos, en la formulación del problema y en la evaluación. En los datos, se incorporarán covariables climáticas con retardos de uno a cinco meses, como temperatura y precipitación de ERA5 [24] y CHIRPS [25] y el índice oceánico de El Niño; índices entomológicos de ovitrampas y encuestas larvarias; y flujos de movilidad humana derivados del transporte o de telefonía móvil agregada, cuyo valor para anticipar la expansión geográfica y el momento de las epidemias está documentado [26]. En la formulación, el paso de la escala anual y nacional a la semanal y subnacional multiplicará las observaciones disponibles y acortará el horizonte, donde el aprendizaje automático ha mostrado ventaja [12]; además, los modelos estimarán la desviación respecto de la persistencia y no el nivel absoluto, de modo que la línea base quede incorporada en la propia predicción. Los pronósticos individuales se combinarán en ensambles ponderados por su desempeño reciente, estrategia que mejoró la calibración en Vietnam [21] y que el ejercicio brasileño de 2024 adoptó como producto final [23]. En la evaluación, la división única se sustituirá por validación con origen móvil y métricas probabilísticas como el WIS, más adecuadas para decidir bajo incertidumbre.

La integración de datos comunitarios constituye un componente innovador y todavía poco aprovechado en la región. Mosquito Alert ha demostrado que los reportes ciudadanos validados por expertos pueden alimentar modelos de aprendizaje profundo capaces de identificar Aedes albopictus con desempeño cercano al humano [14]. Trasladar esta lógica a América Latina, con apoyo institucional de los Ministerios de Salud y los gobiernos locales, podría reducir los retrasos estructurales de la vigilancia pasiva y aumentar la cobertura territorial en áreas rurales y peri-urbanas con baja capacidad entomológica. Plataformas operativas existentes como Infodengue ofrecen una referencia metodológica probada para la integración semanal de fuentes heterogéneas [6], [7].

La participación ciudadana, sin embargo, no puede darse por supuesta. La estrategia prevista combina el codiseño de la aplicación con organizaciones comunitarias, centros educativos y promotores de salud; canales de baja exigencia tecnológica, como mensajería instantánea y formularios que operan sin conexión; y la devolución periódica de mapas de riesgo y de las acciones de control derivadas de cada reporte, pues la retroalimentación visible es el incentivo más directo para sostener el compromiso. Como la brecha digital puede sesgar la cobertura hacia zonas urbanas conectadas, la participación se monitorizará por territorio y se complementará con búsqueda activa en las áreas subrepresentadas. En protección de datos, la plataforma aplicará minimización, sin nombres ni identificadores personales en los reportes; consentimiento informado digital; seudonimización; reducción de la precisión geográfica antes de cualquier publicación; cifrado en tránsito y en reposo; control de acceso por perfiles; y aprobación previa de un comité de ética. Ante la heterogeneidad de los marcos legales nacionales, estos requisitos operarán como estándar mínimo común, sin perjuicio de normas locales más exigentes. En cuanto a recursos, la fase retrospectiva es viable con infraestructura universitaria y software abierto; las fases prospectiva y operativa requieren un equipo mínimo de epidemiología, ciencia de datos, desarrollo de software y coordinación comunitaria, cuyo financiamiento se buscará en convocatorias competitivas nacionales, en la cooperación técnica de la OPS y en fondos internacionales de clima y salud. No existen todavía acuerdos formales con socios institucionales; su establecimiento es condición de entrada a la fase 2.

Las fortalezas del estudio incluyen el uso transparente y verificable de datos abiertos, la división temporal estricta para evitar fuga de información, el reporte explícito de métricas en escala original y logarítmica, y la disponibilidad pública del código de análisis y de las figuras. Las limitaciones son varias. La unidad nacional de análisis oculta la heterogeneidad subnacional crítica para vigilancia operativa. El extracto nacional V1.3 no contiene desagregación por serotipo, gravedad ni grupos de edad, variables esenciales para evaluar carga clínica. Los modelos predictivos no incorporaron variables exógenas climáticas ni de movilidad humana, que la literatura señala como decisivas [10], [13]. por ello no superaron a la persistencia en un año extremo, carencia que las estrategias descritas buscan corregir. La calidad de los datos reportados varía entre países y a lo largo del tiempo, con potencial sub-notificación en territorios con menor capacidad diagnóstica [8]. Finalmente, la propuesta arquitectónica es conceptual y requiere validación piloto con socios institucionales, conforme al protocolo por fases de la sección III-G.

Las líneas futuras incluyen la expansión a la resolución subnacional usando el extracto Spatial_extract_V1_3.csv, para identificar focos locales y heterogeneidades territoriales, la integración de variables climáticas mediante ERA5 y CHIRPS [24],[25], entomológicas y de movilidad, la incorporación de serotipos circulantes, casos de dengue grave y defunciones procedentes de PLISA [15] para caracterizar la carga clínica, el acoplamiento con Infodengue y el dashboard global de la OMS, la colaboración con plataformas de ciencia ciudadana consolidadas, como Mosquito Alert, iNaturalist y GLOBE Observer, para ampliar la cobertura y reducir los retrasos de notificación, la ejecución del piloto descrito en Honduras, Nicaragua y Paraguay; el desarrollo de un módulo de pronóstico probabilístico semanal con horizonte de cuatro semanas similar al ejercicio Infodengue-Mosqlimate. La adopción del marco One Health ampliará el modelo hacia fiebre amarilla, chikungunya, Zika y Oropouche, virus emergente con presión creciente en la cuenca amazónica [5], [11]. La gobernanza de la plataforma deberá alinearse con el Reglamento Sanitario Internacional y con los estándares emergentes de soberanía de datos [22].

V. Conclusiones

Los datos abiertos de OpenDengue V1.3 permiten reconstruir empíricamente la mayor epidemia de dengue jamás documentada en América Latina y el Caribe, con 13.0 millones de casos en 2024 según el análisis presentado. La intensificación regional fue heterogénea pero sincronizada: las razones de casos 2024 frente a la media 2019-2023 alcanzaron o superaron el factor 3 en nueve de los doce países analizados en detalle, con picos de incidencia por 100,000 habitantes en Guyana, Guayana Francesa, Brasil, Paraguay y Honduras. Los modelos de aprendizaje automático evaluados aportaron señales útiles para la clasificación binaria de brotes, con AUC de 0.702, pero no superaron a la línea base de persistencia en regresión durante un año extremo, lo que indica la necesidad de incorporar variables exógenas climáticas, entomológicas, de movilidad y comunitarias, junto con resolución semanal y subnacional y ensambles probabilísticos.

La arquitectura propuesta articula datos oficiales abiertos, datos comunitarios, modelos predictivos reentrenables y mecanismos de gobernanza alineados con el Reglamento Sanitario Internacional. Su valor radica en demostrar que iniciativas regionales como las que el congreso impulsa pueden anclarse en datos verificables, código reproducible y herramientas accesibles desde universidades como la Universidad Nacional Autónoma de Honduras. El siguiente paso es la validación piloto escalonada con socios institucionales en Centroamérica, donde los indicadores de incidencia 2024 indican una urgencia particular en Honduras, Guatemala, Nicaragua y Panamá. Aunque el prototipo se validó con dengue, su diseño permite extenderlo a otras arbovirosis de relevancia regional.

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] S. Bhatt et al., "The global distribution and burden of dengue", *Nature*, vol. 496, no. 7446, pp. 504–507, Apr. 2013, doi: 10.1038/nature12060.

[2] World Health Organization, "Dengue," Aug. 21, 2025. Accessed: May 7, 2026. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue. Accessed: May 7, 2026.

[3] Pan American Health Organization, *Epidemiological Alert: Increase in Dengue Cases in the Americas Region — 7 October 2024*. Washington, DC, USA: PAHO/WHO, 2024. [Online]. Available: https://www.paho.org/en/documents/epidemiological-alert-increase-dengue-cases-americas-region-7-october-2024. Accessed: May 7, 2026.

[4] Pan American Health Organization, "PAHO calls for collective action in response to record increase in dengue cases in the Americas," Mar. 28, 2024. Accessed: May 7, 2026. [Online] Available: https://www.paho.org/en/news/28-3-2024-paho-calls-collective-action-response-record-increase-dengue-cases-americas. Accessed: May 7, 2026.

[5] Pan American Health Organization, "PAHO highlights increase in dengue, Oropouche, and avian influenza cases in the Americas, and advises control measures", *PAHO News Release*, Dec. 10, 2024. [Online]. Available: https://www.paho.org/en/news/10-12-2024-paho-highlights-increase-dengue-oropouche-and-avian-influenza-cases-americas-and. Accessed: May 7, 2026.

[6] C. T. Codeço, F. C. Coelho, O. G. Cruz, S. Oliveira, T. Castro, and L. S. Bastos, "Infodengue: A nowcasting system for the surveillance of arboviruses in Brazil", *Rev. Epidemiol. Sante Publique*, vol. 66, Suppl. 5, p. S386, Jul. 2018, doi: 10.1016/j.respe.2018.05.408.

[7] C. T. Codeço et al., "Fast expansion of dengue in Brazil", *Lancet Reg. Health Am.*, vol. 12, Art. no. 100274, Aug. 2022, doi: 10.1016/j.lana.2022.100274.

[8] J. Clarke, A. Lim, P. Gupte, D. M. Pigott, W. G. van Panhuis, and O. J. Brady, "A global dataset of publicly available dengue case count data", *Sci. Data*, vol. 11, Art. no. 296, 2024, doi: 10.1038/s41597-024-03120-7.

[9] J. Clarke *et al.*, *OpenDengue: Data from the OpenDengue Database*, ver. 1.3, Figshare, 2025, doi: 10.6084/m9.figshare.24259573.

[10] A. Sebastianelli et al., "A reproducible ensemble machine learning approach to forecast dengue outbreaks", *Sci. Rep.*, vol. 14, Art. no. 3807, 2024, doi: 10.1038/s41598-024-52796-9.

[11] M. Cabrera, J. Leake, J. Naranjo-Torres, N. Valero, J. C. Cabrera, and A. J. Rodríguez-Morales, "Dengue prediction in Latin America using machine learning and the One Health perspective: A literature review", *Trop. Med. Infect. Dis.*, vol. 7, no. 10, Art. no. 322, 2022, doi: 10.3390/tropicalmed7100322.

[12] J. M. F. da Costa et al., "Forecasting and early warning systems for dengue outbreaks: Updated narrative review", *Rev. Soc. Bras. Med. Trop.*, vol. 59, Art. no. e0429-2025, 2026, doi: 10.1590/0037-8682-0429-2025.

[13] R. Lowe et al., "Combined effects of hydrometeorological hazards and urbanisation on dengue risk in Brazil: A spatiotemporal modelling study", *Lancet Planet. Health*, vol. 5, no. 4, pp. e209–e219, Apr. 2021, doi: 10.1016/S2542-5196(20)30292-8.

[14] B. A. Pataki, J. Garriga, R. Eritja, J. R. B. Palmer, F. Bartumeus, and I. Csabai, "Deep learning identification for citizen science surveillance of tiger mosquitoes", *Sci. Rep.*, vol. 11, Art. no. 4718, 2021, doi: 10.1038/s41598-021-83657-4.

[15] an American Health Organization, "Dengue indicators," *PAHO/EIH Open Data*. Accessed: May 7, 2026. [Online]. Available: portal actual de Dengue Indicators de PAHO/EIH. PAHO/EIH Dengue Indicators.

[16] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.

[17]M. D. Wilkinson et al., “The FAIR Guiding Principles for scientific data management and stewardship”, *Sci. Data*, vol. 3, Art. no. 160018, 2016, doi: 10.1038/sdata.2016.18.
[18]Pan American Health Organization, “In record year of dengue cases, PAHO urges countries to strengthen response as seasonal transmission set to begin in South America”, *PAHO News Release*, Oct. 8, 2024. [Online]. Available: https://www.paho.org/en/news/8-10-2024-record-year-dengue-cases-paho-urges-countries-strengthen-response-seasonal. Accessed: May 7, 2026.
[19]Pan American Health Organization, *Report on the Epidemiological Situation of Dengue in the Americas: As of Epidemiological Week 12, 2026*. Washington, DC, USA: PAHO/WHO, Apr. 21, 2026. [Online]. Available: documento oficial PAHO. Accessed: May 7, 2026.
[20]K. Heath, L. M. Alves, and M. B. Bonsall, “Climate change, urbanisation and transmission potential: *Aedes aegypti* mosquito projections forecast future arboviral disease hotspots in Brazil”, *PLOS Negl. Trop. Dis.*, vol. 19, no. 9, Art. no. e0013415, 2025, doi: 10.1371/journal.pntd.0013415.
[21]F. J. Colón-González et al., “Probabilistic seasonal dengue forecasting in Vietnam: A modelling study using superensembles”, *PLOS Med.*, vol. 18, no. 3, Art. no. e1003542, Mar. 2021, doi: 10.1371/journal.pmed.1003542.
[22]World Health Assembly, “Revision of the International Health Regulations,” Resolution WHA58.3, World Health Organization, Geneva, Switzerland, May 23, 2005. [Online]. Available: URL oficial WHO. Accessed: May 7, 2026.
[23]E. C. Araujo et al., “Leveraging probabilistic forecasts for dengue preparedness and control: The 2024 Dengue Forecasting Sprint in Brazil”, *Proc. Natl. Acad. Sci. U.S.A.*, vol. 123, no. 7, Art. no. e2508989123, 2026, doi: 10.1073/pnas.2508989123.
[24]H. Hersbach et al., “The ERA5 global reanalysis”, *Q. J. R. Meteorol. Soc.*, vol. 146, no. 730, pp. 1999–2049, 2020, doi: 10.1002/qj.3803.
[25]C. Funk et al., “The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes”, *Sci. Data*, vol. 2, Art. no. 150066, 2015, doi: 10.1038/sdata.2015.66.
[26]A. Wesolowski et al., “Impact of human mobility on the emergence of dengue epidemics in Pakistan”, *Proc. Natl. Acad. Sci. U.S.A.*, vol. 112, no. 38, pp. 11887–11892, 2015, doi: 10.1073/pnas.1504964112.