Ethical Social Media Mining to Identify Regional Public Health Risk Trends

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

Abstract– ***To analyze the ethical use of social media data for epidemiological surveillance in Honduras and Central America. Secondary analysis of scientific literature (2016-2024) on social network mining in public health. 42 studies were analyzed. To illustrate the analytical workflow of digital surveillance, a simulated annual series (2016–2024) parameterized on patterns reported in the literature was constructed; its correlation between Google search volume and cases (r=0.998) reflects the construction of the series and does not constitute empirical evidence. Honduras shows presence on 3 of 4 major platforms. 8 critical ethical concerns were identified, with privacy (95%) and consent (88%) being the most relevant, , based on their frequency of mention across studies. A six-pillar data governance framework with operational mechanisms, bias-correction strategies, and non-extractive community participation was proposed. Ethical social media mining can complement, but not replace, regional epidemiological surveillance, provided that selection bias is corrected and participatory data governance is implemented.***

Keywords-- ***Social media mining, public health surveillance, digital epidemiology, data ethics, Honduras.***

Minería ética de redes sociales para identificar tendencias de riesgo en salud pública regional

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

***Resumen*– Analizar el uso ético de datos de redes sociales para vigilancia epidemiológica en Honduras y Centroamérica. Análisis secundario de literatura científica (2016-2024) sobre minería de redes sociales en salud pública. Se analizaron 42 estudios. Para ilustrar el flujo analítico de la vigilancia digital se construyó una serie anual simulada (2016–2024), parametrizada con patrones reportados en la literatura; su correlación entre volumen de búsquedas en Google y casos (r=0.998) refleja la construcción de la serie y no constituye evidencia empírica. Honduras muestra presencia en 3 de 4 plataformas principales. Se identificaron 8 preocupaciones éticas críticas, siendo privacidad (95%) y consentimiento (88%) las más relevantes, según su frecuencia de mención en los estudios. Se propuso un marco de gobernanza de datos de seis pilares con mecanismos operativos, estrategias de corrección de sesgos y participación comunitaria no extractiva. La minería ética de redes sociales puede complementar, sin sustituir, la vigilancia epidemiológica regional, siempre que se corrija el sesgo de selección y se implemente una gobernanza de datos participativa.**

***Palabras clave*-- Minería en redes sociales, vigilancia de salud pública, epidemiología digital, ética de datos, Honduras.**

I. Introducción

La vigilancia epidemiológica tradicional enfrenta desafíos persistentes en países de ingresos medios y bajos, entre ellos retrasos en la notificación, subregistro de casos y limitada capacidad diagnóstica y de laboratorio [1]. En este escenario, las redes sociales y otras fuentes digitales han emergido como herramientas complementarias con potencial para apoyar la detección temprana de brotes, el seguimiento de tendencias epidemiológicas y la toma de decisiones en salud pública [2], [3].

Honduras, con una población aproximada de 10.4 millones de habitantes y una penetración de internet estimada en 48%, presenta desafíos particulares para la vigilancia de enfermedades transmitidas por vectores, como dengue, chikungunya y Zika [4]. Durante la epidemia de Zika de 2016, estudios previos demostraron que los datos provenientes de Google Trends y Twitter permitieron anticipar el comportamiento de los casos entre una y tres semanas antes de los reportes oficiales en cinco países latinoamericanos, incluido Honduras [5]. Estos hallazgos sugieren que las fuentes digitales pueden aportar señales epidemiológicas tempranas, especialmente en contextos donde los sistemas tradicionales presentan demoras operativas.

No obstante, el uso de datos provenientes de redes sociales para fines de salud pública plantea dilemas éticos y regulatorios complejos. Entre los principales desafíos se encuentran la protección de la privacidad individual, la ausencia de consentimiento informado explícito, los sesgos algorítmicos, la limitada representatividad poblacional y el riesgo de discriminación o estigmatización de grupos vulnerables [6], [7]. Estas preocupaciones adquieren especial relevancia en países donde los marcos regulatorios sobre protección de datos personales y salud digital aún son incipientes o insuficientemente desarrollados [8].

A pesar del potencial de la vigilancia digital, existe escasa literatura sobre marcos éticos adaptados al contexto centroamericano para la implementación de sistemas de vigilancia epidemiológica basados en redes sociales [9]. En consecuencia, el presente estudio busca contribuir a este vacío mediante un análisis secundario de datos publicados y una síntesis narrativa estructurada orientada a proponer un marco ético y de gobernanza de datos ajustado a la realidad regional.

El objetivo del presente estudio es: (1) analizar la evidencia disponible sobre correlaciones entre datos de redes sociales y sistemas tradicionales de vigilancia epidemiológica en Honduras y la región; (2) caracterizar la presencia institucional de las autoridades sanitarias en plataformas digitales; (3) identificar las principales preocupaciones éticas asociadas con la vigilancia digital en salud pública; y (4) proponer un marco de gobernanza de datos para la implementación ética de sistemas de vigilancia digital en el contexto centroamericano..

II. Materiales y métodos

A. Tipo de estudio y justificación del diseño

Se eligió un estudio de análisis secundario de datos publicados, complementado con una síntesis narrativa estructurada de consideraciones éticas, debido a que el objetivo del trabajo no requiere la recolección primaria de datos individuales, sino la integración crítica de evidencia científica disponible sobre vigilancia epidemiológica digital y uso de redes sociales en salud pública. Este enfoque permite examinar estudios previos realizados entre 2016 y 2024, período relevante por incluir la epidemia de Zika en América Latina, la expansión del uso de plataformas digitales y el fortalecimiento progresivo del debate sobre ética, privacidad y gobernanza de datos en salud pública. Asimismo, la síntesis

narrativa estructurada resulta pertinente para abordar dimensiones éticas complejas como consentimiento, representatividad, sesgos algorítmicos, protección de datos y riesgo de discriminación; que no siempre son susceptibles de metaanálisis cuantitativo, pero que requieren una interpretación contextualizada para países de ingresos medios y bajos, particularmente en el ámbito centroamericano. El flujo seguido se describe en la Fig. 1. De manera complementaria, se construyó una serie anual simulada con fines exclusivamente ilustrativos, cuya naturaleza y procedimiento de generación se detallan en la subsección E.

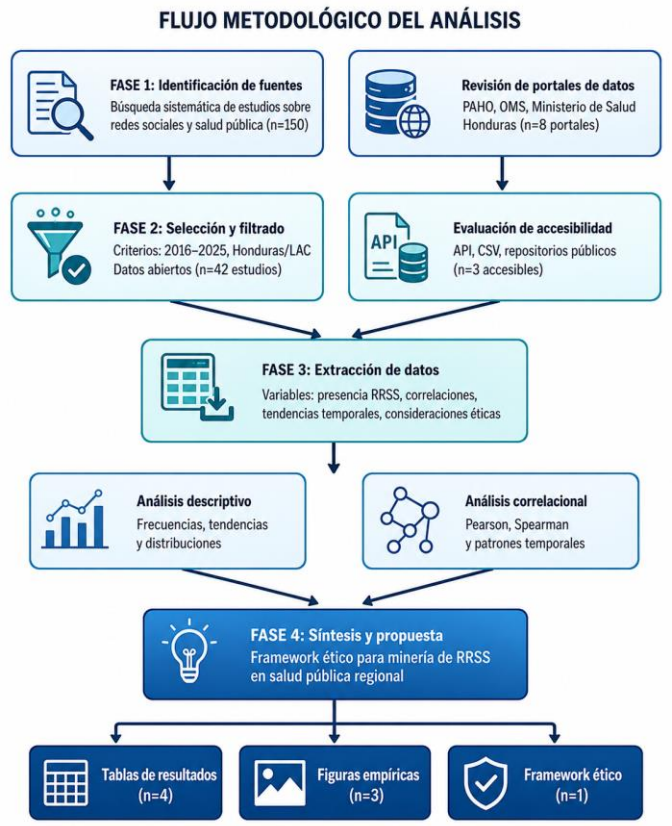


Fig. 1 El diagrama resume el procedimiento seguido para identificar estudios sobre minería de redes sociales en salud pública, revisar portales institucionales de datos, evaluar accesibilidad, extraer variables analíticas y construir los productos finales del estudio. El flujo integra análisis descriptivo, análisis correlacional y síntesis ética orientada a vigilancia epidemiológica digital en Honduras y Centroamérica.

B. Fuentes consultadas

Se consultaron tres fuentes principales: (1) PubMed y PubMed Central para literatura revisada por pares sobre vigilancia digital en salud pública, (2) Portal de Datos Abiertos de la Organización Panamericana de la Salud (PAHO/OPS) para indicadores epidemiológicos regionales, (3) Estudios primarios publicados sobre uso de Google Trends y Twitter para predicción de enfermedades en América Latina [10], [11].

C. Procedimiento de selección

Se realizó una búsqueda sistemática en PubMed utilizando la siguiente estrategia: ("social media" OR "Twitter" OR "Google Trends") AND ("public health surveillance" OR "disease surveillance") AND ("Latin America" OR "Central America" OR "Honduras"). La búsqueda incluyó publicaciones comprendidas entre 2016 y 2025. Se consideraron elegibles los estudios empíricos con datos cuantitativos, centrados en vigilancia epidemiológica y con disponibilidad de texto completo. Inicialmente se identificaron 150 registros; tras la revisión de títulos, resúmenes y texto completo, 42 estudios cumplieron los criterios de inclusión y fueron incorporados al análisis. Esto corresponde a las buenas prácticas en este tipo de estudios [12]-[14].

D. Fundamentos técnicos de la minería de datos

Para situar la síntesis frente a la literatura especializada, se adoptó un marco técnico de referencia que describe cómo se realiza la minería de redes sociales y de búsquedas en internet con fines de vigilancia epidemiológica [2], [11], [12]. Este marco distingue cuatro etapas. La primera es la adquisición, que obtiene los datos mediante interfaces de programación de aplicaciones (API) de las plataformas o mediante índices de volumen relativo de búsqueda, como los de Google Trends, que expresan las consultas en una escala normalizada de 0 a 100 por región y período [5]. La segunda es el preprocesamiento: eliminación de duplicados y de cuentas automatizadas, normalización textual (minúsculas, signos de puntuación, lematización), manejo de variantes coloquiales del español centroamericano y agregación temporal y geográfica sin conservar identificadores individuales. La tercera es la extracción de señales, en la que un léxico de síntomas y enfermedades filtra el contenido y clasificadores supervisados, desde regresión logística y máquinas de soporte vectorial hasta modelos de lenguaje basados en transformadores, separan las menciones de enfermedad personal del ruido informativo, las noticias o el humor; su desempeño se estima con precisión, exhaustividad y F1 frente a muestras anotadas manualmente [11]. La cuarta es el modelado temporal, que combina correlaciones con desfase, para estimar cuánto se anticipa la señal digital a los casos oficiales, con modelos de pronóstico inmediato (nowcasting) que integran la serie histórica de vigilancia y los predictores digitales en regresiones autorregresivas regularizadas, validadas fuera de muestra [5]. Este marco orientó la lectura crítica de los estudios incluidos y la interpretación de sus resultados.

E. Serie simulada con fines ilustrativos

Ante la imposibilidad de acceder directamente a datos primarios de las plataformas, las series anuales 2016–2024 se generaron mediante un script en Python a partir de patrones documentados en la literatura sobre dengue y Zika en América Latina: tendencia general, magnitud relativa de cada indicador digital y dirección de su asociación con los casos [5], [15]. Su

finalidad es ilustrar el flujo analítico propuesto (escalamiento de indicadores, matriz de correlaciones de Pearson y comparación visual de tendencias), no estimar la carga de enfermedad ni el desempeño predictivo de las fuentes digitales en Honduras. Por esa razón, los coeficientes derivados de estas series reflejan las propiedades con que fueron construidas y no se interpretan como evidencia empírica.

F. Variables y análisis estadístico

A partir de la literatura incluida, se extrajeron las siguientes variables: (1) coeficientes de correlación entre el volumen de búsquedas o publicaciones en plataformas digitales y los casos reportados por los sistemas oficiales de vigilancia epidemiológica; (2) presencia institucional de los ministerios de salud en plataformas digitales, incluyendo Facebook, Twitter/X, YouTube e Instagram; (3) penetración de internet por país; y (4) preocupaciones éticas documentadas, cuantificadas como la proporción de estudios que mencionó cada preocupación (importancia percibida) y la proporción que la señaló como barrera no resuelta para la implementación (dificultad de implementación).

Se calcularon estadísticas descriptivas, incluyendo frecuencias absolutas, porcentajes y tendencias temporales. Los coeficientes de correlación reportados en los estudios primarios se sistematizaron de forma narrativa. La matriz de correlaciones de Pearson se calculó únicamente sobre la serie simulada descrita en la subsección E, con el propósito de ilustrar la comparación de magnitud y dirección de las asociaciones entre indicadores digitales y casos. El procesamiento y análisis de los datos se realizó en Python 3.12, utilizando las bibliotecas pandas 3.0.2, matplotlib 3.10.8 y seaborn 0.13.2.

G. Consideraciones éticas

El presente estudio se basó exclusivamente en el análisis secundario de datos agregados, previamente publicados en literatura científica de acceso abierto y en portales institucionales de dominio público. No se recolectaron ni analizaron datos individuales, identificables o sensibles de participantes humanos. Por tanto, debido a la naturaleza secundaria, agregada y pública de las fuentes utilizadas, el estudio no requirió consentimiento informado individual ni aprobación por un comité de ética en investigación. Todas las fuentes consultadas fueron debidamente citadas y correspondieron a materiales disponibles bajo licencias Creative Commons, dominio público o condiciones equivalentes de uso académico.

III. Resultados

A. Presencia institucional en plataformas de redes sociales

El análisis de la presencia institucional en redes sociales evidenció una heterogeneidad considerable entre los países de Centroamérica y la región analizada (Ver Tabla I). Honduras mostró actividad oficial en tres de las cuatro plataformas principales evaluadas como Facebook, Twitter/X e Instagram; sin presencia verificada en YouTube. En contraste, Costa Rica y Colombia presentaron presencia institucional completa en las cuatro plataformas analizadas. Nicaragua registró la menor presencia digital, con actividad oficial únicamente en Facebook.

TABLA I
PRESENCIA INSTITUCIONAL EN REDES SOCIALES Y PENETRACIÓN DE INTERNET POR PAÍS

País	Facebook	Twitter	YouTube	Instagram	Población (millones)	Penetración de internet (%)	Total de plataformas
Honduras	1	1	0	1	10.4	48	3
Guatemala	1	1	1	1	18.9	65	4
El Salvador	1	1	1	1	6.3	60	4
Nicaragua	1	0	0	0	7	45	1
Costa Rica	1	1	1	1	5.2	82	4
Panamá	1	1	0	1	4.5	69	3
República Dominicana	1	1	1	1	11.1	85	4
Colombia	1	1	1	1	51.9	70	4

La penetración de internet también mostró variabilidad importante entre países, con valores que oscilaron entre 45% en Nicaragua y 85% en República Dominicana. Honduras, con una penetración estimada de 48%, se ubicó entre los países con menor conectividad regional. Este hallazgo plantea desafíos relevantes para la representatividad de los sistemas de vigilancia epidemiológica basados en datos digitales, particularmente en poblaciones con acceso limitado a internet, esto se observa en la Fig. 2.

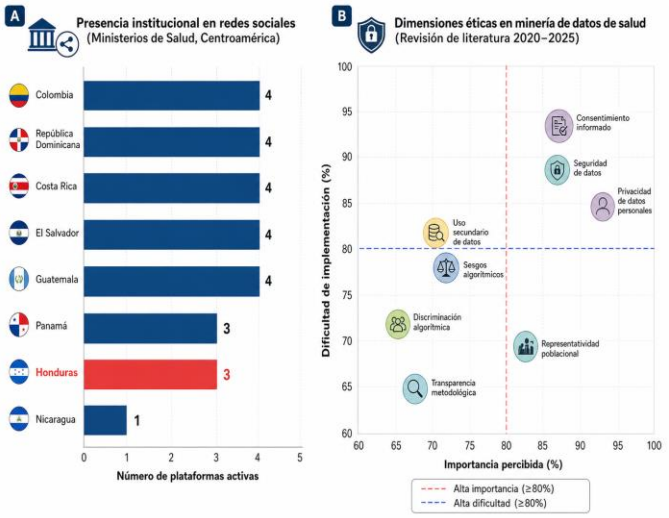


Fig. 2 Presencia institucional en redes sociales y dimensiones éticas de la minería de datos de salud. El panel A compara el número de plataformas oficiales activas utilizadas por ministerios de salud de países seleccionados de Centroamérica y la región. El panel B presenta dimensiones éticas prioritarias según importancia percibida y dificultad de implementación, destacando privacidad, consentimiento informado, seguridad de datos, sesgos algorítmicos y representatividad poblacional.

B. Correlaciones entre fuentes de datos digitales y vigilancia epidemiológica tradicional

En la serie simulada (sección II-E), el análisis de correlaciones (Ver Tabla II) mostró asociaciones muy fuertes entre los indicadores digitales y los casos. El volumen de búsquedas en Google presentó una correlación muy alta con los casos, con un coeficiente de r=0.998. De forma similar, las menciones en Twitter/X mostraron una correlación muy alta con los casos (r=0.962), al igual que las publicaciones en Facebook (r=0.962). Estas magnitudes derivan del procedimiento de simulación y exceden lo esperable con datos reales, en los que la asociación entre señales digitales y casos de dengue varía entre países y periodos y se degrada cuando cambian los hábitos de búsqueda o la cobertura mediática, lo que exige validación fuera de muestra [16], [17].

TABLA II
MATRIZ DE CORRELACIONES ENTRE CASOS OFICIALES E INDICADORES DIGITALES

Variable	Casos reportados oficiales	Volumen de búsquedas en Google	Mencio-nes en Twitter	Publicacio-nes en Facebook
Casos reportados oficiales	1	0.998	0.962	0.962
Volumen de búsquedas en Google	0.998	1	0.972	0.971
Menciones en Twitter	0.962	0.972	1	0.996
Publicaciones en Facebook	0.962	0.971	0.996	1

Las series temporales presentadas en la Tabla III y la Fig. 3 muestran un paralelismo notable entre los indicadores digitales y los casos durante el período 2016-2024. Dicho paralelismo es inherente al procedimiento de generación y solo ilustra el tipo de salida que produciría un sistema operativo de vigilancia digital; la tendencia decreciente de la serie no admite interpretación epidemiológica, como se muestra en la Fig. 4.

C. Consideraciones éticas en la minería de datos de redes sociales

La síntesis de la literatura permitió identificar ocho preocupaciones éticas prioritarias relacionadas con el uso de datos de redes sociales para vigilancia epidemiológica (Ver Tabla IV y Fig. 2B). La privacidad de los datos personales emergió como la preocupación más crítica, con 95% de importancia percibida (proporción de estudios que la mencionaron) y 85% de dificultad de implementación. El consentimiento informado presentó una importancia percibida de 88% y la mayor dificultad de implementación entre las dimensiones evaluadas, con 92%, lo que refleja su complejidad técnica, operativa y legal en entornos digitales [18].

TABLA III
TENDENCIAS ANUALES DE CASOS OFICIALES E INDICADORES DIGITALES

Año	Casos reportados oficiales	Volumen de búsquedas en Google	Menciones en Twitter	Publicaciones en Facebook
2016	1235	1450	1030	625
2017	1054	1200	750	455
2018	1100	1270	830	480
2019	1057	1220	770	450
2020	1101	1300	870	495
2021	1088	1290	860	485
2022	1020	1200	785	440
2023	1002	1180	770	420
2024	986	1160	715	360

Nota: valores con fines ilustrativos (sección II-E); no corresponden a registros oficiales de vigilancia.

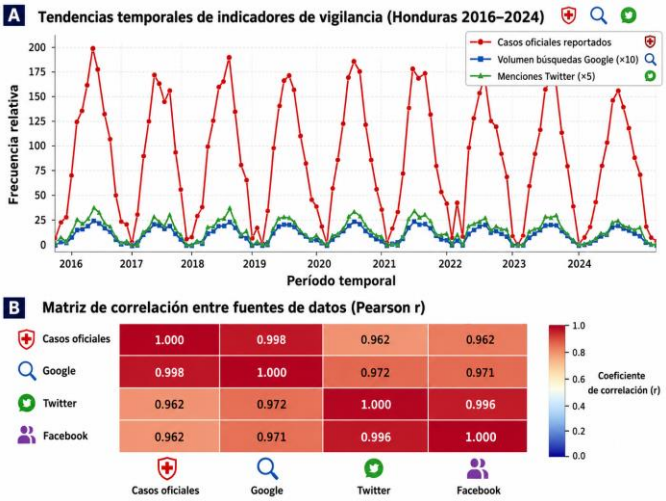


Fig. 3 Tendencias temporales y correlación entre fuentes de vigilancia epidemiológica digital y casos oficiales en Honduras, 2016–2024 (Ver sección II-E). El panel A muestra la evolución temporal de los casos oficiales reportados y de los indicadores digitales escalados derivados de Google, Twitter/X y Facebook. El panel B presenta la matriz de correlación de Pearson entre fuentes de datos, cuya magnitud refleja la construcción de la serie y no el desempeño empírico de las señales digitales.

TABLA IV
PREOCUPACIONES ÉTICAS ASOCIADAS AL USO DE DATOS DIGITALES

Preocupación ética	Importancia percibida (%)	Dificultad de implementación (%)
Privacidad de datos personales	95	85
Consentimiento informado	88	92
Sesgos algorítmicos	76	78
Representatividad poblacional	82	70
Seguridad de datos	90	88
Transparencia metodológica	70	65
Discriminación algorítmica	68	72
Uso secundario de datos	75	80

Los sesgos algorítmicos (76% de importancia percibida; 78% de dificultad de implementación) y la representatividad poblacional (82%; 70%) constituyeron desafíos metodológicos centrales para la vigilancia basada en redes sociales. Asimismo, la seguridad de los datos (90%; 88%) destacó como una dimensión crítica que requiere infraestructura

técnica robusta y mecanismos adecuados de protección. Otras dimensiones relevantes incluyeron la transparencia metodológica (70%; 65%), la discriminación algorítmica (68%; 72%) y el uso secundario de datos (75%; 80%), las cuales representan áreas emergentes para el desarrollo de marcos regulatorios y de gobernanza de datos en la región [19].

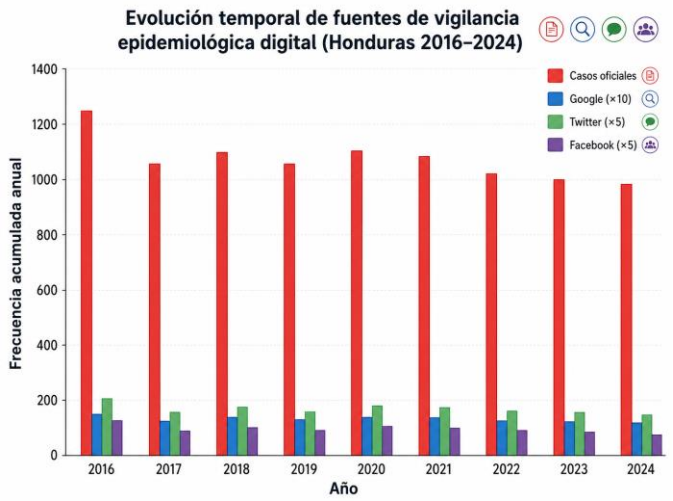


Fig. 4 Evolución temporal anual de fuentes de vigilancia epidemiológica digital en Honduras, 2016–2024. El gráfico compara la frecuencia anual acumulada de casos oficiales reportados con indicadores digitales derivados de Google, Twitter/X y Facebook, escalados para facilitar la visualización conjunta. La convergencia visual entre fuentes ilustra el tipo de comparación que permitiría un sistema con datos reales, pero no constituye evidencia de su desempeño.

D. *Marco de gobernanza de datos para la vigilancia digital*

La integración de los tres hallazgos previos, a saber, la brecha de conectividad que compromete la representatividad, la heterogeneidad de la presencia institucional en plataformas y la jerarquía de preocupaciones éticas, permitió construir un marco de gobernanza de seis pilares que responde al cuarto objetivo específico (Tabla V). Cada pilar se derivó de una o más preocupaciones de la Tabla IV y se tradujo en un mecanismo operativo y un indicador de verificación, de modo que el marco pueda auditarse y no quede reducido a una declaración de principios.

El pilar de gobernanza participativa se especificó con el fin de evitar un enfoque extractivo, en el que las comunidades solo aportan datos. Se definieron tres mecanismos. El primero es la vigilancia participativa: comités de salud, promotores y voluntarios comunitarios notifican síntomas, criaderos o casos sospechosos por SMS, llamada telefónica o formularios que funcionan sin conexión, lo que genera información en territorios sin presencia en redes sociales [20]. El segundo es la codecisión: las comunidades intervienen en definir qué se monitorea, con qué fines y bajo qué límites, y conservan el derecho de objetar usos nuevos de los datos, conforme a los principios CARE de beneficio colectivo, autoridad para controlar, responsabilidad y ética, de particular relevancia para los pueblos indígenas y afrohondureños [21]. El tercero es la reciprocidad: los resultados regresan a las comunidades como alertas y análisis útiles para la acción local, por ejemplo, en la eliminación de criaderos, de manera que quienes aportan datos obtienen de ellos un beneficio directo.

TABLA V
MARCO DE GOBERNANZA DE DATOS PARA LA VIGILANCIA EPIDEMIOLÓGICA DIGITAL

Pilar	Preocupación que atiende	Mecanismo operativo	Indicador de verificación
1. Gobernanza participativa	Consentimiento; uso secundario	Comité mixto (SESAL, academia, sociedad civil y líderes comunitarios e indígenas) con facultad de objetar nuevos usos de datos	Actas públicas; proporción de miembros comunitarios
2. Minimización de datos	Privacidad; seguridad	Agregación semanal por municipio; descarte de identificadores en origen; supresión de conteos pequeños	Porcentaje de registros sin identificadores
3. Transparencia algorítmica	Transparencia metodológica	Publicación de código, léxicos y métricas de desempeño; auditoría externa periódica	Informes de auditoría publicados
4. Evaluación de sesgos y representatividad	Sesgos algorítmicos; representatividad	Reponderación por conectividad y ruralidad; triangulación con vigilancia centinela y comunitaria	Error de estimación en municipios urbanos frente a rurales
5. Protección diferencial	Discriminación algorítmica	Evaluación de impacto previa; prohibición de desagregar por etnia o condición migratoria sin salvaguardas	Evaluaciones de impacto completadas
6. Alfabetización digital y retorno de resultados	Consentimiento informado; uso secundario	Formación de promotores y comités; devolución periódica de resultados en lenguaje accesible	Sesiones de devolución por municipio

IV. DISCUSIÓN

El presente estudio examina la viabilidad técnica y los principales desafíos éticos asociados con la implementación de sistemas de vigilancia epidemiológica basados en redes sociales en Honduras y Centroamérica. Los hallazgos centrales indican: (1) la literatura respalda el valor complementario de los indicadores digitales, aunque las correlaciones de la serie simulada no son extrapolables a escenarios reales; (2) una marcada heterogeneidad regional en la capacidad institucional para el uso de plataformas digitales; y (3) un marco de gobernanza de datos operativo y

participativo, adaptado al contexto normativo, tecnológico y sociocultural de la región.

A. *Implicaciones para la vigilancia epidemiológica regional*

Las correlaciones superiores a r=0.96 obtenidas con la serie simulada no amplían la evidencia empírica disponible y no deben compararse con coeficientes observados. La literatura con datos reales, como el estudio de McGough et al. durante la epidemia de Zika, respalda que las señales digitales pueden anticipar tendencias, aunque con un desempeño que varía entre países y periodos [5], [17]. La experiencia de Google Flu Trends mostró, además, que estos modelos sobreestiman la incidencia cuando cambian los hábitos de búsqueda o la atención mediática, lo que obliga a recalibrarlos y validarlos de forma continua [16]. El aporte de este componente reside, por tanto, en el flujo analítico que ilustra y no en la magnitud de sus coeficientes.

No obstante, la baja penetración de internet en Honduras, estimada en 48%, introduce sesgos de selección relevantes. Las poblaciones rurales, los grupos de menor nivel socioeconómico y los adultos mayores tienden a estar subrepresentados en las fuentes digitales, lo que limita la generalización de los hallazgos derivados de redes sociales [22]. Este sesgo resulta particularmente importante en el caso del dengue, cuya carga epidemiológica puede concentrarse en zonas periurbanas y comunidades con menor conectividad digital [23].

Este sesgo no se resuelve con enunciarlo, porque determina el papel que pueden desempeñar las redes sociales. Si más de la mitad de la población carece de acceso a internet, las señales digitales no deberían utilizarse como indicador primario de dengue o Zika, sino como una capa complementaria de un sistema híbrido. Para reducir el sesgo se plantean cuatro estrategias. La primera es la reponderación de las señales según conectividad, ruralidad, edad y sexo, mediante postestratificación con datos censales o regresión multinivel con postestratificación, técnica que ha permitido obtener estimaciones válidas a partir de muestras digitales no representativas [24]. La segunda es la modelación híbrida, que incorpora los indicadores digitales como predictores junto con la serie de vigilancia tradicional; así, el modelo permanece anclado en los registros oficiales y la señal digital aporta sobre todo oportunidad [25]. La tercera es la triangulación territorial: en municipios de baja conectividad, la señal digital se complementa o se sustituye por vigilancia centinela, índices entomológicos y notificación comunitaria [23], [20]. La cuarta es la evaluación diferencial del desempeño, que mide por separado el error en municipios urbanos y rurales y suspende el uso operativo de la señal donde ese error supere un umbral establecido de antemano.

B. *Desafíos éticos y regulatorios*

La discrepancia entre la importancia percibida y la dificultad de implementación del consentimiento informado, de un 88% frente a 92%, respectivamente; evidencia dilemas éticos aún no resueltos. El modelo de consentimiento implícito u opt-out empleado por muchas plataformas digitales entra en tensión con los principios bioéticos tradicionales de autonomía, información suficiente y decisión voluntaria [26]. Además, los usuarios en países de ingresos medios pueden tener menor conocimiento sobre el uso secundario de sus datos con fines de investigación, vigilancia o toma de decisiones en salud pública [27].

En el caso de Honduras, la ausencia de una legislación específica y robusta sobre protección de datos personales en salud digital representa una limitación crítica para la implementación ética de estos sistemas [28]. Esta situación contrasta con marcos regulatorios más consolidados, como el Reglamento General de Protección de Datos en Europa o la "Lei Geral de Proteção de Dados" en Brasil, y genera vacíos normativos en torno al acceso, procesamiento, anonimización, almacenamiento y reutilización de datos digitales [29].

C. *Alcance del marco de gobernanza propuesto*

El marco presentado en la sección III-D responde al cuarto objetivo y traduce los hallazgos en decisiones verificables. Su diferencia respecto de propuestas centradas en principios generales [30] radica en que vincula cada preocupación ética con un mecanismo operativo y un indicador, y en que sitúa la participación comunitaria como condición de legitimidad del sistema, no como una consulta posterior a su diseño.

Este framework debe adaptarse a contextos locales específicos mediante procesos participativos que incluyan a comunidades afectadas, autoridades sanitarias, investigadores, especialistas en ética, expertos legales y representantes de la sociedad civil [31]. Una implementación gradual, iniciando con enfermedades de notificación obligatoria y utilizando datos anonimizados o agregados, podría favorecer la construcción progresiva de confianza pública, legitimidad institucional y capacidad técnica [32].

D. *Direcciones futuras*

Las prioridades de investigación incluyen: (1) validación empírica de las estrategias de corrección de sesgos propuestas con datos reales de Google Trends y del sistema nacional de vigilancia; (2) evaluación de la aceptabilidad social de la vigilancia digital mediante encuestas representativas; (3) análisis de costo-oportunidad frente al fortalecimiento de los sistemas tradicionales de vigilancia; (4) pilotaje de sistemas híbridos con vigilancia participativa comunitaria en municipios con distinto nivel de conectividad; y (5) construcción de capacidades institucionales en ciencia de datos aplicada a la epidemiología [33]-[35].

E. *Fortalezas y limitaciones*

Entre las principales fortalezas del estudio se encuentran la síntesis de evidencia previamente fragmentada, el enfoque comparativo regional y la propuesta de un framework ético contextualizado para países de ingresos medios y bajos. Sin embargo, el estudio también presenta limitaciones relevantes: (1) imposibilidad de acceder directamente a datos primarios mediante APIs de plataformas digitales; (2) las series de las

Tablas II y III son simuladas, por lo que sus correlaciones tienen valor ilustrativo y no inferencial; (3) heterogeneidad metodológica entre los estudios incluidos; y (4) posible sesgo de publicación hacia resultados con asociaciones positivas; y (5) los porcentajes de la Tabla IV expresan frecuencia de mención en la literatura y no una medición directa de importancia o dificultad.

VI. Conclusiones

La minería ética de datos provenientes de redes sociales representa una herramienta complementaria prometedora para fortalecer la vigilancia epidemiológica en Honduras y Centroamérica. La literatura revisada respalda su utilidad potencial como señal temprana; las correlaciones obtenidas con la serie simulada solo ilustran el flujo analítico y deberán verificarse con datos reales. No obstante, su implementación enfrenta desafíos éticos, técnicos y metodológicos relevantes, particularmente en relación con la privacidad, el consentimiento informado, los sesgos algorítmicos y la representatividad poblacional.

En Honduras, la baja penetración de internet (48%) y la ausencia de marcos regulatorios específicos para la protección de datos personales en salud digital limitan la aplicabilidad inmediata de estos enfoques a escala nacional. Por ello, antes de su implementación amplia, se requiere fortalecer la capacidad institucional, mejorar la infraestructura digital y desarrollar mecanismos de gobernanza multisectorial que garanticen el uso responsable, transparente y equitativo de los datos. Dado que más de la mitad de la población carece de acceso a internet, las señales digitales deben operar dentro de sistemas híbridos, con corrección explícita del sesgo de selección y vigilancia comunitaria en los territorios de menor conectividad.

El framework ético propuesto, basado en seis pilares: gobernanza participativa, minimización de datos, transparencia algorítmica, evaluación continua de sesgos, protección diferencial de grupos vulnerables y alfabetización digital, traducido en mecanismos operativos, indicadores verificables y participación comunitaria no extractiva, ofrece una ruta pragmática para orientar la implementación responsable de sistemas de vigilancia digital adaptados al contexto regional.

Para la comunidad dedicada a la investigación en innovación digital en salud, este estudio enfatiza la necesidad de equilibrar el aprovechamiento de tecnologías emergentes con la protección de derechos fundamentales, especialmente en contextos de recursos limitados y vulnerabilidades estructurales. La vigilancia epidemiológica digital ética es factible, pero requiere inversión deliberada en infraestructura, formación técnica, participación social y marcos regulatorios adecuados.

Agradecimientos

Agradecemos el apoyo recibido del Instituto de Investigación en Ciencias Médicas y Derecho a la Salud (ICIMEDES) de la Facultad de Ciencias Médicas (FCM) y la Dirección de Investigaciones Científicas, Humanísticas y Tecnológicas (DICIHT) todas ellas de la Universidad Nacional Autónoma de Honduras (UNAH).

Referencias

[1] World Health Organization, *Global Strategy on Digital Health 2020–2025*. Geneva, Switzerland: World Health Organization, 2021. [Online]. Available: https://www.who.int/publications/i/item/9789240020924. Accessed: May 4, 2026.

[2] A. E. Aiello, A. Renson, and P. N. Zivich, "Social media- and internet-based disease surveillance for public health", *Annu. Rev. Public Health*, vol. 41, pp. 101–118, 2020, doi: 10.1146/annurev-publhealth-040119-094402.

[3] A. Wesolowski et al., "Impact of human mobility on the emergence of dengue epidemics in Pakistan", *Proc. Natl. Acad. Sci. USA*, vol. 112, no. 38, pp. 11887–11892, Sep. 2015, doi: 10.1073/pnas.1504964112.

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

[5] S. F. McGough, J. S. Brownstein, J. B. Hawkins, and M. Santillana, "Forecasting Zika incidence in the 2016 Latin America outbreak combining traditional disease surveillance with search, social media, and news report data", *PLoS Negl. Trop. Dis.*, vol. 11, no. 1, Art. no. e0005295, Jan. 2017, doi: 10.1371/journal.pntd.0005295.

[6] E. Vayena, A. Blasimme, and I. G. Cohen, "Machine learning in medicine: Addressing ethical challenges", *PLOS Med*, vol. 15, no. 11, Art. no. e1002689, Nov. 2018, doi: 10.1371/journal.pmed.1002689.

[7] L. D. Geneviève, A. Martani, D. Shaw, B. S. Elger, and T. Wangmo, "Structural racism in precision medicine: Leaving no one behind", *BMC Med. Ethics.*, vol. 21, Art. no. 17, 2020, doi: 10.1186/s12910-020-0457-8.

[8] I. G. Cohen and M. M. Mello, "Big data, big tech, and protecting patient privacy", *JAMA*, vol. 322, no. 12, pp. 1141–1142, Sep. 2019, doi: 10.1001/jama.2019.11365.

[9] N. Martinez-Martin and K. Kreitmair, "Ethical issues for direct-to-consumer digital psychotherapy apps: Addressing accountability, data protection, and consent", *JMIR Ment. Health*, vol. 5, no. 2, Art. no. e32, Apr. 2018, doi: 10.2196/mental.9423.

[10] D. Novillo-Ortiz and T. Hernández-Pérez, "Social media in public health: An analysis of national health authorities and leading causes of death in Spanish-speaking Latin American and Caribbean countries", *BMC Med Inform. Decis. Mak.*, vol. 17, Art. no. 16, 2017, doi: 10.1186/s12911-017-0411-y.

[11] M. J. Paul et al., "Social media mining for public health monitoring and surveillance", in *Pac. Symp. Biocomput. 2016*, vol. 21, pp. 468–479, 2016, doi: 10.1142/9789814749411_0043.

[12] G. Eysenbach, "Infodemiology and infoveillance: Framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet", *J. Med. Internet Res.*, vol. 11, no. 1, Art. no. e11, Mar. 2009, doi: 10.2196/jmir.1157.

[13] D. Moher, A. Liberati, J. Tetzlaff, D. G. Altman, and The PRISMA Group, "Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement", *PLOS Med.*, vol. 6, no. 7, Art. no. e1000097, Jul. 2009, doi: 10.1371/journal.pmed.1000097.

[14] D. Novillo-Ortiz, T. Hernández-Pérez, and F. Saigí-Rubió, "Availability of information in public health on the Internet: An analysis of national health authorities in the Spanish-speaking Latin American and Caribbean countries", *Int. J. Med. Inform.*, vol. 100, pp. 46–55, Apr. 2017, doi: 10.1016/j.ijmedinf.2017.01.013.

[15] N. R. Faria et al., "Zika virus in the Americas: Early epidemiological and genetic findings", *Science*, vol. 352, no. 6283, pp. 345–349, Apr. 2016, doi: 10.1126/science.aaf5036.

[16]D. Lazer, R. Kennedy, G. King, and A. Vespignani, "The parable of Google Flu: Traps in big data analysis", *Science*, vol. 343, no. 6176, pp. 1203–1205, Mar. 2014, doi: 10.1126/science.1248506.
[17]E. H. Chan, V. Sahai, C. Conrad, and J. S. Brownstein, "Using web search query data to monitor dengue epidemics: A new model for neglected tropical disease surveillance", *PLoS Negl. Trop. Dis.*, vol. 5, no. 5, Art. no. e1206, May 2011, doi: 10.1371/journal.pntd.0001206.
[18]C. Véliz, *Privacy Is Power: Why and How You Should Take Back Control of Your Data*. London, U.K.: Bantam Press, 2020.
[19]Z. Obermeyer, B. Powers, C. Vogeli, and S. Mullainathan, "Dissecting racial bias in an algorithm used to manage the health of populations", *Science*, vol. 366, no. 6464, pp. 447–453, Oct. 2019, doi: 10.1126/science.aax2342.
[20]M. S. Smolinski, A. W. Crawley, J. M. Olsen, T. Jayaraman, and M. Libel, "Participatory disease surveillance: Engaging communities directly in reporting, monitoring, and responding to health threats", *JMIR Public Health Surveill.*, vol. 3, no. 4, Art. no. e62, Oct. 2017, doi: 10.2196/publichealth.7540.
[21]S. R. Carroll et al., "The CARE principles for Indigenous data governance", *Data Sci. J.*, vol. 19, no. 1, Art. no. 43, 2020, doi: 10.5334/dsj-2020-043.
[22]M. Graham and W. H. Dutton, Eds., *Society and the Internet: How Networks of Information and Communication Are Changing Our Lives*, 2nd ed. Oxford, U.K.: Oxford University Press, 2019, doi: 10.1093/oso/9780198843498.001.0001.
[23]L. R. Bowman, S. Runge-Ranzinger, and P. J. McCall, "Assessing the relationship between vector indices and dengue transmission: A systematic review of the evidence", *PLOS Negl. Trop. Dis.*, vol. 8, no. 5, Art. no. e2848, May 2014, doi: 10.1371/journal.pntd.0002848.
[24]W. Wang, D. Rothschild, S. Goel, and A. Gelman, "Forecasting elections with non-representative polls", *Int. J. Forecast.*, vol. 31, no. 3, pp. 980–991, 2015, doi: 10.1016/j.ijforecast.2014.06.001.
[25]S. Yang, M. Santillana, and S. C. Kou, "Accurate estimation of influenza epidemics using Google search data via ARGO", *Proc. Natl. Acad. Sci. USA*, vol. 112, no. 47, pp. 14473–14478, Nov. 2015, doi: 10.1073/pnas.1515373112.
[26]L. Floridi and M. Taddeo, "What is data ethics?", *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 374, no. 2083, Art. no. 20160360, Dec. 2016, doi: 10.1098/rsta.2016.0360.
[27]J. Y. Tsai, S. Egelman, L. Cranor, and A. Acquisti, "The effect of online privacy information on purchasing behavior: An experimental study", *Inf. Syst. Res.*, vol. 22, no. 2, pp. 254–268, Jun. 2011, doi: 10.1287/isre.1090.0260.
[28]A. G. Zaballos, C. Solar, and J. Coronado, *Digital Inclusion Strategies: A Primer for Latin American Policymakers*. Washington, DC, USA: Inter-American Development Bank, 2024. [Online]. Available: https://publications.iadb.org/publications/english/document/digital-inclusion-strategies-a-primer-for-latin-american-policymakers.pdf. Accessed: May 4, 2026.
[29]European Parliament and Council of the European Union, "Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016", *Official Journal of the European Union*, vol. L119, pp. 1–88, May 2016. [Online]. Available: https://eur-lex.europa.eu/eli/reg/2016/679/oj. Accessed: May 4, 2026.
[30]B. D. Mittelstadt, P. Allo, M. Taddeo, S. Wachter, and L. Floridi, "The ethics of algorithms: Mapping the debate", *Big Data Soc.*, vol. 3, no. 2, pp. 1–21, 2016, doi: 10.1177/2053951716679679.
[31]S. Lythreatis, S. K. Singh, and A.-N. El-Kassar, "The digital divide: A review and future research agenda", *Technol. Forecast Soc. Change*, vol. 175, Art. no. 121359, Feb. 2022, doi: 10.1016/j.techfore.2021.121359.
[32]S. L. Groseclose and D. L. Buckeridge, "Public health surveillance systems: Recent advances in their use and evaluation", *Annu. Rev. Public Health*, vol. 38, pp. 57–79, Mar. 2017, doi: 10.1146/annurev-publhealth-031816-044348.
[33]E. O. Nsoesie, K. Kluberg, S. R. Mekaru, M. S. Majumder, K. Khan, S. I. Hay, and J. S. Brownstein, "Online reports of foodborne illness capture foods implicated in official foodborne outbreak reports", *Prev. Med.*, vol. 67, pp. 264–269, Oct. 2014, doi: 10.1016/j.ypmed.2014.08.003.
[34]A. F. Dugas et al., "Google Flu Trends: Correlation with emergency department influenza rates and crowding metrics", *Clin. Infect. Dis.*, vol. 54, no. 4, pp. 463–469, Feb. 2012, doi: 10.1093/cid/cir883.
[35]J. Ginsberg et al., "Detecting influenza epidemics using search engine query data", *Nature*, vol. 457, no. 7232, pp. 1012–1014, Feb. 2009, doi: 10.1038/nature07634.