

Detección temprana de brotes de enfermedades infecciosas mediante análisis de datos de redes sociales y fuentes digitales: una revisión de alcance

Yamil Liscano Martínez

Asesor

Luis Angel Anillo Arrieta

Universidad Nacional Abierta y a Distancia UNAD
Escuela de Ciencias Básicas, Tecnología e Ingeniería – ECBTI
Especialización Ciencia de Datos y Analítica

2025

Resumen

Contexto/Objetivo: La vigilancia digital, basada en el análisis de fuentes digitales como redes sociales y motores de búsqueda, ha emergido como un método prometedor para detectar tempranamente brotes de enfermedades infecciosas. Esta revisión sistemática tuvo como objetivo evaluar integralmente las metodologías, desempeño, y limitaciones de estas técnicas digitales frente a los métodos tradicionales de vigilancia epidemiológica. **Metodología:** Se siguieron las directrices del Joanna Briggs Institute y PRISMA-SCR, utilizando bases de datos científicas (PubMed, Scopus, Web of Science, entre otras). Se incluyeron estudios empíricos y revisiones sistemáticas sobre vigilancia digital de enfermedades infecciosas, sin restricción de idioma, analizando fuentes digitales, algoritmos analíticos, precisión y validación con datos oficiales. **Resultados:** La vigilancia digital logra anticipar brotes desde días hasta semanas antes que métodos tradicionales, con alta precisión (coeficientes de correlación >0.8 y bajos errores predictivos). Predominan fuentes como Google Trends y Twitter, aplicando técnicas como regresiones supervisadas, algoritmos bayesianos y modelos ARIMA. Sin embargo, la calidad y representatividad de los datos varían según resolución espacial, selección de términos clave, ruido mediático y factores demográficos. **Conclusiones:** Pese a su capacidad predictiva demostrada, se recomienda estandarizar protocolos, incorporar análisis demográficos y aprovechar inteligencia artificial para reducir sesgos, optimizando así la vigilancia epidemiológica digital.

Palabras claves: Brote de Enfermedades; Vigilancia Epidemiológica; Redes Sociales; Infodemiología; Inteligencia Artificial; Análisis de Series Temporales.

Abstract

Context/Objective: Digital surveillance, based on the analysis of digital sources such as social media and search engines, has emerged as a promising method for the early detection of infectious disease outbreaks. This systematic review aimed to comprehensively assess the methodologies, performance, and limitations of these digital techniques compared to traditional epidemiological surveillance methods. **Methodology:** Guidelines from the Joanna Briggs Institute and PRISMA-SCR were followed, using scientific databases (PubMed, Scopus, Web of Science, among others). Empirical studies and systematic reviews on digital surveillance of infectious diseases were included, with no language restriction, analyzing digital sources, analytical algorithms, accuracy, and validation with official data. **Results:** Digital surveillance can anticipate outbreaks from days to weeks before traditional methods, with high accuracy (correlation coefficients >0.8 and low predictive errors). Predominant sources include Google Trends and Twitter, applying techniques such as supervised regression, Bayesian algorithms, and ARIMA models. However, the quality and representativeness of the data vary depending on spatial resolution, selection of key terms, media noise, and demographic factors. **Conclusions:** Despite its demonstrated predictive capacity, it is recommended to standardize protocols, incorporate demographic analysis, and leverage artificial intelligence to reduce biases, thus optimizing digital epidemiological surveillance.

Keywords: Disease Outbreaks; Epidemiological Surveillance; Social Media; Infodemiology; Artificial Intelligence; Time Series Analysis

Tabla de contenido

Introducción	9
Descripción del Problema	12
Planteamiento del Problema	13
Sistematización del Problema	15
Justificación	16
Objetivos	18
Objetivo General	18
Objetivos Específicos.....	18
Estado del Arte.....	19
Panorama Actual de la Investigación y Tendencias	19
Metodologías Predominantes y Emergentes	20
Avances Recientes y Tecnologías Clave	21
Marco Contextual.....	24
Marco Teórico.....	25
Fundamentos de Epidemiología y Vigilancia	25
Teorías de Difusión de Información Digital	25
Principios de Ciencia de Datos Aplicados	26
Informática en Salud Pública	27
Marco Conceptual.....	28
Detección Temprana y Brote	28
Datos y Análisis	28
Campos Relacionados.....	29

Marco Normativo y Consideraciones Éticas.....	30
Directrices Éticas de la OMS (2017)	30
Regulaciones de Privacidad (Ej. HIPAA en EE.UU.)	31
Técnicas de Anonimización y Desidentificación.....	31
Consentimiento, Confianza y Participación Comunitaria.....	32
Metodología	33
Protocolo	33
Criterios de Elegibilidad	33
Pregunta PCC (Población, Concepto, Contexto)	34
Estrategia de Búsqueda	34
Selección de Estudios y Extracción de Datos	34
Análisis Estadístico	36
Resultados.....	37
Información General del Estudio	37
Fuentes de Datos y Plataformas Digitales	43
Métodos y Técnicas Analíticas	48
Desempeño y Detección Temprana	55
Aspectos Complementarios y Específicos	61
Discusión.....	68
Principales Hallazgos y Evolución Metodológica	68
Integración de Fuentes Digitales con Sistemas Tradicionales.....	68
Capacidades de Detección Temprana y Precisión Predictiva.....	69

Consideraciones Complementarias: Resolución, Selección de Palabras Clave, Impacto Mediático y Variables Demográficas	69
Comparación con la Literatura Previa	70
Consideraciones Metodológicas	76
Discusión de Técnicas Específicas	76
Claridad en el Proceso de Integración	77
Selección de Palabras Clave y Gestión del Impacto Mediático.....	78
Limitaciones de los Estudios	78
Implicaciones Clínicas y Recomendaciones para Investigaciones Futuras	80
Recomendaciones Claras y Accionables	80
Perspectiva Interdisciplinaria.....	82
Conclusiones	83
Bibliografía	84

Lista de Tablas

Tabla 1 <i>Características Generales de los Estudios Incluidos</i>	39
Tabla 2 <i>Fuentes de Datos Digitales y Plataformas Empleadas</i>	43
Tabla 3 <i>Métodos de Análisis y Procesamiento de Datos</i>	48
Tabla 4 <i>Desempeño en Detección Temprana y Precisión</i>	56
Tabla 5 <i>Características Complementarias y Variables Contextuales</i>	61
Tabla 6 <i>Comparación de Resultados con Otros Estudios</i>	71

Lista de Figuras

Figura 1 <i>Diagrama PRISMA</i>	38
Figura 2 <i>Distribución Temporal y Geográfica de los Estudios</i>	42
Figura 3 <i>Frecuencia de Uso de Diferentes Fuentes de Datos</i>	46
Figura 4 <i>Diagrama de Flujo Metodológico</i>	53
Figura 5 <i>Matriz de Técnicas Analíticas en Estudios de Vigilancia Digital</i>	55
Figura 6 <i>Comparación del Lead Time y Precisión entre Estudios</i>	59
Figura 7 <i>Relación Entre la Ventaja Temporal y la Tasa de Detección en Estudios de Vigilancia</i>	60
Figura 8 <i>Mapa de Calor de las Características Complementarias</i>	67

Introducción

Las enfermedades infecciosas emergentes constituyen una amenaza creciente para la biodiversidad y las sociedades humanas, generando impactos significativos tanto en ecosistemas como en la salud pública global. Tradicionalmente, la vigilancia epidemiológica ha dependido principalmente de sistemas oficiales que, en ocasiones, enfrentan retrasos en la detección temprana de brotes. En este contexto, la vigilancia digital ha surgido como una herramienta complementaria que permite captar señales tempranas en tiempo real, utilizando datos provenientes de fuentes no convencionales (Fallatah & Adekola, 2024; MacIntyre et al., 2023; O'Shea, 2017; Shausan et al., 2023a).

Plataformas digitales como motores de búsqueda, redes sociales y sistemas de vigilancia participativa se han convertido en herramientas clave para captar indicios tempranos de actividad epidémica. Diversos estudios han demostrado, por ejemplo, que señales obtenidas de búsquedas en Baidu o publicaciones en Twitter pueden anticipar la detección de enfermedades como COVID-19 e influenza entre una y tres semanas antes de los reportes oficiales de los sistemas tradicionales (Dai & Wang, 2020; Kogan et al., 2021; Perlaza et al., 2024; Shin et al., 2016).

La integración de múltiples fuentes digitales, incluyendo consultas en Google, mensajes de Twitter y alertas en plataformas como HealthMap, permite lograr altos niveles de correlación con datos oficiales, además de mejorar significativamente la capacidad predictiva. Técnicas avanzadas como algoritmos de aprendizaje automático, modelos autoregresivos y métodos estadísticos de suavizado temporal han demostrado incrementar la precisión y proporcionar alertas con suficiente anticipación (Poirel et al., 2019; Samaras et al., 2020; Shausan et al., 2023a).

Estos métodos no se restringen a un único contexto geográfico; su efectividad se ha validado en múltiples países, incluyendo China, Reino Unido, Estados Unidos y diversos países

de América Latina. La alta resolución espacial y temporal de estos datos digitales facilita la identificación precisa de focos locales de transmisión, lo que permite acciones específicas de prevención y control (Budd et al., 2020; Chen et al., 2022; Dhewantara et al., 2019; Nageshwaran et al., 2021; Villanueva Parra et al., 2024).

Sin embargo, la vigilancia digital enfrenta ciertos desafíos relacionados con la calidad de los datos, como la presencia de ruido mediático, sesgos por disparidades en el acceso a internet y la necesidad imperiosa de integrar eficazmente estos métodos con los sistemas tradicionales de vigilancia epidemiológica. Para asegurar la confiabilidad de estas señales, son esenciales el filtrado minucioso de los datos, la selección cuidadosa de palabras clave y la validación rigurosa a través de análisis estadísticos como correlaciones, RMSE y curvas ROC (Aiello et al., 2020; Fallatah & Adekola, 2024; Ibrahim, 2020; Zhao et al., 2021).

Adicionalmente, la vigilancia participativa y el análisis de contenido en redes sociales han permitido el desarrollo de sistemas eficientes de alerta temprana. Estudios recientes han logrado identificar brotes de COVID-19 hasta 16 días antes que los sistemas oficiales, así como detectar oportunamente epidemias de MRSA y dengue con niveles altos de sensibilidad y especificidad (Araújo et al., 2020; Dueñas et al., 2023; Hussain-Alkhateeb et al., 2018; Meckawy et al., 2022).

La evidencia acumulada indica que, mediante el adecuado preprocesamiento de datos digitales y la aplicación de técnicas analíticas avanzadas, es posible mejorar significativamente la respuesta frente a brotes infecciosos. La integración efectiva de señales digitales con métodos tradicionales representa, por tanto, una estrategia crucial para optimizar la vigilancia epidemiológica y facilitar la toma rápida de decisiones en salud pública (Fallatah & Adekola, 2024; Shausan et al., 2023a).

El objetivo de esta revisión es evaluar integralmente los diferentes enfoques de vigilancia digital utilizados para el monitoreo y detección temprana de brotes infecciosos. Se revisarán en detalle los métodos de preprocesamiento de datos, los algoritmos analíticos empleados, las métricas de desempeño utilizadas y las estrategias de integración de estas señales digitales con los sistemas tradicionales de salud pública, identificando fortalezas, limitaciones y posibles áreas de mejora.

Descripción del Problema

La amenaza constante que representan las enfermedades infecciosas para la salud global se ha visto exacerbada en las últimas décadas por factores como el cambio climático, el aumento de la densidad poblacional, la globalización y la rápida movilidad humana (Dueñas et al., 2023; Groseclose & Buckeridge, 2017). La detección temprana de brotes es un pilar fundamental para mitigar su impacto, permitiendo la implementación oportuna de medidas de control y prevención que salvan vidas y reducen las cargas económicas y sociales. Sin embargo, los sistemas tradicionales de vigilancia epidemiológica, basados principalmente en la notificación pasiva de casos por parte de instituciones de salud, a menudo enfrentan limitaciones inherentes relacionadas con retrasos en la recopilación, consolidación y análisis de datos (Ibrahim, 2020; O'Shea, 2017). Estos retrasos pueden generar una ventana crítica de tiempo durante la cual un brote puede expandirse significativamente antes de ser detectado oficialmente, como se evidenció en las fases iniciales de la pandemia de COVID-19 (Budd et al., 2020; Ibrahim, 2020). La necesidad de cerrar esta brecha temporal e identificar señales de alerta con mayor anticipación ha impulsado la búsqueda de herramientas y enfoques complementarios que aprovechen las vastas cantidades de datos generados en tiempo real a través de plataformas digitales.

Planteamiento del Problema

Ante los desafíos que presentan los sistemas de vigilancia epidemiológica convencionales para la detección temprana de brotes de enfermedades infecciosas y el creciente volumen de información digital disponible, emerge la vigilancia digital como una herramienta prometedora (Aiello et al., 2020; Groseclose & Buckeridge, 2017). Diversos estudios han explorado el uso de datos provenientes de redes sociales, motores de búsqueda y otras fuentes digitales para captar indicios tempranos de actividad epidémica, demostrando en algunos casos una capacidad significativa de anticipación. Sin embargo, la aplicación de la vigilancia digital en salud pública se caracteriza por una notable heterogeneidad en las fuentes de datos utilizadas (como Twitter, Google Trends, noticias online, etc.), las metodologías de análisis empleadas (que van desde análisis de correlación simples hasta modelos complejos de aprendizaje automático) y las métricas de desempeño reportadas (Al-Kenane et al., 2024; Groseclose & Buckeridge, 2017; Sharpe et al., 2016). Existe una necesidad clara de comprender de manera sistemática y exhaustiva el estado actual del conocimiento sobre cómo se están utilizando estas fuentes digitales, qué técnicas analíticas resultan más efectivas, cuál es el rendimiento real en términos de detección temprana y precisión, y cuáles son las principales limitaciones y desafíos que impiden su plena integración y estandarización con los sistemas de vigilancia tradicionales. Sin una comprensión clara de estos aspectos, resulta difícil optimizar el potencial de la vigilancia digital, seleccionar las herramientas adecuadas para contextos específicos y garantizar la confiabilidad y representatividad de las señales de alerta generadas (O'Shea, 2017; Timpka et al., 2014; Zhao et al., 2021).

En este contexto, el problema de investigación que aborda esta revisión es la falta de una síntesis integral y estructurada de la literatura científica que describa y evalúe las metodologías,

el desempeño y las limitaciones de la detección temprana de brotes de enfermedades infecciosas mediante el análisis de datos de redes sociales y fuentes digitales, así como el grado de integración de estos enfoques con los sistemas de vigilancia epidemiológica tradicionales.

Sistematización del Problema

Para abordar el problema de investigación planteado, la presente revisión de alcance busca responder a las siguientes preguntas específicas:

¿Cuáles son las principales fuentes de datos digitales y plataformas utilizadas en los estudios de vigilancia epidemiológica para la detección temprana de brotes de enfermedades infecciosas?

¿Qué metodologías de preprocesamiento de datos y técnicas analíticas (algoritmos estadísticos y de aprendizaje automático) se emplean en la detección de señales epidemiológicas a partir de fuentes digitales?

¿Cómo se evalúa el desempeño de los sistemas de vigilancia digital, y cuáles son los resultados reportados respecto a la ventaja temporal en la detección, la tasa de detección y la precisión predictiva?

¿Cuáles son las estrategias y el grado de integración reportado entre los métodos de vigilancia digital y los sistemas de vigilancia de salud pública tradicionales?

¿Cuáles son las principales limitaciones y desafíos asociados al uso de fuentes digitales para la vigilancia epidemiológica, incluyendo aspectos de resolución espacial y temporal, proceso de selección de palabras clave, medición del impacto mediático y consideraciones demográficas o de uso?

Justificación

La emergencia y reemergencia de enfermedades infecciosas representan una amenaza constante y creciente para la salud pública a nivel global. La rapidez con la que estas enfermedades pueden propagarse, como lo ha evidenciado la pandemia de COVID-19, subraya la necesidad crítica de contar con sistemas de vigilancia epidemiológica ágiles y eficientes que permitan la detección temprana de brotes. Tradicionalmente, la vigilancia se ha basado en reportes de casos confirmados por laboratorios y sistemas de salud, los cuales, si bien son fundamentales, a menudo presentan un retraso inherente en la notificación, limitando la capacidad de respuesta oportuna de las autoridades sanitarias (McClymont et al., 2024; Shausan et al., 2023b).

En este contexto, el auge de las tecnologías digitales y el uso masivo de plataformas como las redes sociales y los motores de búsqueda han generado un vasto y dinámico repositorio de datos que reflejan las inquietudes, síntomas y comportamientos de la población en tiempo real. Este fenómeno ha dado origen a la vigilancia digital, un enfoque innovador que busca aprovechar estas fuentes de datos no convencionales para identificar señales tempranas de actividad epidémica, potencialmente antes de que los casos sean oficialmente reportados a través de los canales tradicionales (Fallatah & Adekola, 2024; Seo Dong-Woo, 2017).

Si bien diversos estudios preliminares y experiencias recientes sugieren el enorme potencial de la vigilancia digital para complementar y fortalecer los sistemas existentes, la heterogeneidad de las fuentes de datos utilizadas, la variedad de metodologías analíticas aplicadas y las diferencias en el desempeño y la validación de estos sistemas hacen necesaria una evaluación integral y sistemática. Comprender a fondo las metodologías empleadas, las fuentes digitales más efectivas, el rendimiento real en términos de anticipación y precisión, así como las

limitaciones intrínsecas de estos enfoques, es crucial para su implementación efectiva y su integración en las estrategias de salud pública (Groseclose & Buckeridge, 2017; Shakeri Hossein Abad et al., 2021).

Por lo tanto, la presente revisión de alcance se justifica por la necesidad de sintetizar y mapear la literatura existente sobre la detección temprana de brotes de enfermedades infecciosas mediante el análisis de datos de redes sociales y fuentes digitales. Al evaluar de manera comprensiva las metodologías, el desempeño y las limitaciones reportadas en los estudios publicados, esta investigación contribuirá a clarificar el estado actual del conocimiento en este campo emergente. Los hallazgos de esta revisión proporcionarán información valiosa para investigadores, profesionales de la salud pública y tomadores de decisiones, orientando futuras investigaciones, el desarrollo de herramientas de vigilancia digital más robustas y la formulación de estrategias más efectivas para la prevención y el control de brotes, sentando así las bases para una respuesta epidemiológica más proactiva y eficiente en la era digital.

Objetivos

Objetivo General

Evaluar integralmente las metodologías, desempeño, limitaciones y estrategias de integración de los sistemas de vigilancia digital basados en el análisis de datos de redes sociales y fuentes digitales para la detección temprana de brotes de enfermedades infecciosas.

Objetivos Específicos

Caracterizar las fuentes de datos digitales, plataformas tecnológicas y metodologías empleadas en la vigilancia de enfermedades infecciosas, incluyendo técnicas de preprocesamiento y análisis de datos como algoritmos estadísticos y de aprendizaje automático.

Evaluar el desempeño de los sistemas de vigilancia digital en términos de ventaja temporal en la detección, tasa de detección y precisión predictiva, examinando el grado y las estrategias de integración de sus hallazgos con los sistemas oficiales de salud pública, en comparación con los sistemas de vigilancia tradicionales.

Analizar las limitaciones, desafíos metodológicos y barreras para la implementación reportados en la literatura sobre vigilancia digital de enfermedades infecciosas, incluyendo aspectos como la resolución espacial y temporal, la selección de palabras clave, el impacto mediático y las consideraciones demográficas.

Estado del Arte

El campo de la vigilancia de enfermedades infecciosas utilizando fuentes digitales ha experimentado una rápida evolución, estableciéndose como un complemento vital, aunque complejo, a los sistemas tradicionales de salud pública.

Panorama Actual de la Investigación y Tendencias

La epidemiología digital y la vigilancia digital han surgido como disciplinas que aprovechan la ubicuidad de los datos generados en línea para monitorizar la salud poblacional (Terry et al., 2023). Estos enfoques no buscan reemplazar, sino complementar los sistemas de vigilancia tradicionales, ofreciendo potenciales ventajas en términos de oportunidad y alcance (Terry et al., 2023). La investigación en este ámbito, especialmente la que utiliza datos de redes sociales, se ha consolidado en torno a tres grandes ejes temáticos: la monitorización y vigilancia (incluyendo detección y predicción de enfermedades, seguimiento de actitudes y sentimiento público), la comunicación de riesgos (difusión de consejos de salud, comportamiento de búsqueda de información, circulación de desinformación o "infodemias") y el control de enfermedades (respuesta gubernamental, cambios en el comportamiento público, calidad de la información educativa) (Luo et al., 2025; Shakeri Hossein Abad et al., 2021).

Si bien los primeros estudios exploratorios datan de la primera década del siglo XXI, la investigación sistemática y el volumen de publicaciones comenzaron a crecer notablemente alrededor de 2010. Este crecimiento experimentó una aceleración exponencial con la llegada de la pandemia de COVID-19. La pandemia actuó como un catalizador significativo, impulsando un aumento exponencial en el volumen de investigación en esta área. Esta aceleración se debió a la necesidad urgente de vigilancia en tiempo real y a la prominencia de las redes sociales como canales primarios para el discurso público y la búsqueda de información relacionada con la

pandemia, generando conjuntos de datos masivos y accesibles que atrajeron a los investigadores. Sin embargo, esta misma aceleración podría introducir un sesgo en la literatura existente, con un enfoque predominante en aplicaciones específicas de la pandemia, lo que podría eclipsar las particularidades de otras enfermedades o desafíos metodológicos fundamentales que existían previamente (Al-Kenane et al., 2024; Dueñas et al., 2023; McClymont et al., 2024).

Metodológicamente, el campo ha estado dominado por estudios observacionales, análisis de contenido y enfoques correlacionales. No obstante, se observa una transición gradual. A medida que el campo madura, el enfoque se desplaza desde la simple demostración del *potencial* de los datos digitales hacia investigaciones más predictivas y evaluativas. Existe un interés creciente en validar estos métodos, evaluar su impacto real en la salud pública y compararlos rigurosamente con los sistemas tradicionales y otros benchmarks. La disponibilidad de herramientas analíticas más sofisticadas, como el aprendizaje automático y el aprendizaje profundo, está facilitando la realización de tareas predictivas más complejas, impulsando esta tendencia hacia la evaluación de la utilidad práctica más allá de la exploración académica.

Las fuentes de datos utilizadas son cada vez más diversas, yendo más allá de las redes sociales y las búsquedas web para incluir registros electrónicos de salud (EHRs), datos de movilidad derivados de teléfonos móviles, datos de sensores portátiles (wearables), informes de noticias, e incluso datos de vigilancia de aguas residuales. Esta diversificación refleja un reconocimiento de las limitaciones de cualquier fuente única y apunta hacia el desarrollo de sistemas de vigilancia híbridos o multimodales.

Metodologías Predominantes y Emergentes

La transformación de datos digitales en información epidemiológica útil se basa en una combinación de metodologías avanzadas:

Minería de Datos y PLN: Permiten extraer significado de textos no estructurados (redes sociales, EHRs, noticias), útiles para farmacovigilancia, detección de síntomas, análisis de percepción pública y determinantes sociales de salud. Se emplean desde modelos estadísticos (LDA, ATAM) hasta modelos de aprendizaje profundo.

Aprendizaje Automático y Profundo: Utilizados para clasificación, predicción y detección de brotes. Modelos como GBM, SVM, LSTM y Transformers destacan por su capacidad para modelar patrones complejos y temporales en datos clínicos y epidemiológicos.

Análisis de Series Temporales: Métodos como ARIMA, Prophet y redes neuronales (DeepAR, LSTM) se emplean para predecir tendencias en la incidencia de enfermedades, con resultados variables según el contexto.

Epidemiología Espacial y SIG: Herramientas esenciales para mapear riesgos, visualizar distribución geográfica de casos y optimizar estrategias de vigilancia y muestreo.

Vigilancia Sindrómica: Analiza datos prediagnósticos (síntomas, búsquedas en internet) para detectar anomalías en tiempo real antes de que se confirmen diagnósticos.

Modelos Compartimentales y Basados en Agentes: Simulan la dinámica de transmisión de enfermedades usando datos digitales (movilidad, redes sociales) para ajustar parámetros y hacer predicciones más realistas.

Avances Recientes y Tecnologías Clave

La evolución tecnológica está potenciando de manera significativa la vigilancia digital mediante seis pilares clave:

Inteligencia Artificial y Aprendizaje Profundo: Modelos avanzados como LSTM, GRU, TCN y Transformers mejoran el diagnóstico asistido por IA, el pronóstico clínico y la predicción

de brotes, gracias a su capacidad para modelar relaciones complejas y temporales en datos epidemiológicos.

Plataformas Integradas (como DHIS2): Facilitan la vigilancia en tiempo real mediante la integración de datos clínicos, de laboratorio y administrativos, permitiendo una respuesta rápida y centralizada, como se ha visto en países con sistemas electrónicos robustos.

Tecnologías de Biodetección Avanzadas: Dispositivos como bionanosensores, tiras reactivas y plataformas microfluídicas permiten la detección rápida y en el punto de atención, generando datos inmediatos para alimentar sistemas digitales de vigilancia.

Computación en la Nube y Big Data: Proveen la infraestructura necesaria para almacenar, procesar y analizar grandes volúmenes de datos en tiempo real, esenciales para el entrenamiento de modelos complejos de IA y la gestión de información a gran escala.

Escucha Social e Infodemia: Herramientas como EARS usan IA para monitorear redes sociales, detectar desinformación, identificar narrativas emergentes y apoyar la respuesta sanitaria frente a la sobrecarga informativa (infodemia).

Integración de Datos Multimodales: Se promueve la combinación de fuentes heterogéneas (movilidad, clima, genómica, EHRs, aguas residuales) para superar limitaciones de los datos digitales puros, construyendo sistemas de vigilancia híbridos más precisos y completos.

Lagunas Identificadas en el Conocimiento y Áreas Prioritarias de Investigación

A pesar del progreso, la vigilancia digital enfrenta múltiples desafíos críticos que requieren investigación prioritaria:

Evaluación de Efectividad: Faltan estudios comparativos rigurosos que midan sensibilidad, especificidad y oportunidad real de detección frente a métodos tradicionales. La

ausencia de métricas estandarizadas y la baja calidad de datos en algunas regiones limitan el impacto práctico de estos sistemas.

Sesgos en los Datos Digitales: Existen importantes sesgos (demográficos, algorítmicos, de plataforma), pero hay escasa investigación orientada a cuantificarlos y corregirlos, lo que compromete la validez y generalización de los hallazgos.

Integración y Protocolos Operativos: Se necesitan guías claras y protocolos validados para integrar los sistemas digitales en la rutina de salud pública de manera efectiva y sostenible.

Marco Ético y de Privacidad: Urge un marco ético específico para datos digitales, que aborde temas como consentimiento, anonimización, propiedad de datos y estigmatización, más allá de las directrices generales existentes.

Causalidad vs Correlación: Predomina la investigación observacional, sin establecer vínculos causales sólidos entre indicadores digitales (como búsquedas o tuits) y eventos reales de salud, lo que reduce la confiabilidad para intervenciones predictivas.

Exposición Multiplataforma: Falta estudiar cómo la información (y desinformación) recibida desde múltiples redes influye conjuntamente en comportamientos de salud, ya que la mayoría de estudios se enfocan en una sola plataforma.

Control Activo de Enfermedades: La utilidad práctica de los datos digitales para acciones concretas de control (como rastreo de contactos o promoción de conductas) sigue siendo teórica y necesita validación empírica.

Equidad y Generalizabilidad: La evidencia se concentra en ciertas enfermedades, plataformas y regiones ricas. Es clave ampliar el enfoque a contextos diversos y regiones de bajos ingresos, donde la vigilancia digital podría ser más útil, pero enfrenta mayores limitaciones estructurales.

Marco Contextual

La vigilancia digital de enfermedades se apoya en un ecosistema global diverso de fuentes de datos (redes sociales, motores de búsqueda, medios digitales, registros electrónicos, aplicaciones móviles, wearables y aguas residuales), lo que permite detectar señales de salud en tiempo real. Sin embargo, enfrenta múltiples desafíos: acceso limitado y desigual a los datos, baja calidad y ruido informativo, sesgos diversos (demográficos, de plataforma, metodológicos) y falta de validación clínica directa (Aiello et al., 2020; Terry et al., 2023).

A nivel global, organizaciones como el CDC y la OMS lideran estrategias para estandarizar, capacitar e integrar herramientas digitales en sistemas de salud. Ejemplos exitosos incluyen la vigilancia participativa en África (Sierra Leona, Tanzania), Asia (Tailandia, Camboya), América Latina (modelo ACS en Brasil) y países de altos ingresos (como Reino Unido), donde plataformas adaptadas a contextos locales han demostrado mayor sostenibilidad que aquellas basadas únicamente en fuentes externas (Fallatah & Adekola, 2024).

La integración efectiva con sistemas de salud existentes ofrece oportunidades para complementar y fortalecer la vigilancia tradicional, especialmente mediante modelos híbridos. No obstante, aún persisten barreras como la falta de interoperabilidad, infraestructura limitada, escasez de personal capacitado, problemas de gobernanza de datos y resistencias culturales. Superar estos retos requiere inversiones sostenidas, estandarización técnica y políticas claras que garanticen confianza, privacidad y uso ético de la información.

Marco Teórico

La detección temprana de brotes usando datos digitales se basa en la integración de teorías de varias disciplinas:

Fundamentos de Epidemiología y Vigilancia

La epidemiología aporta los conceptos básicos para entender cómo surgen y se propagan los brotes (cadena de transmisión, historia natural de la enfermedad) (Fallatah & Adekola, 2024; Tarkoma et al., 2020).

La vigilancia en salud pública es la recolección, análisis e interpretación sistemática de datos para guiar la acción. Existen varios modelos:

Pasiva: Recibe reportes (común, bajo costo, pero con subregistro y retrasos).

Activa: Busca casos activamente (más completa, pero costosa).

Centinela: Monitorea en sitios seleccionados (representativa, calidad sobre cantidad).

Sindrómica: Monitorea síntomas pre-diagnósticos (más oportuna, menos específica).

Basada en Eventos (EBS): Detecta eventos en fuentes no estructuradas/no oficiales (noticias, web).

La Teoría de la Vigilancia en Salud Poblacional (PHST) describe la vigilancia como un proceso de 5 pasos y distingue entre monitoreo (recolección de datos crudos) y vigilancia (análisis avanzado para generar conocimiento accionable o "inteligencia epidemiológica"). Las señales digitales a menudo actúan como "desencadenantes" iniciales que requieren esta inteligencia para ser útiles (Fallatah & Adekola, 2024; Tarkoma et al., 2020).

Teorías de Difusión de Información Digital

Es clave entender cómo se propaga la información (correcta o incorrecta) en línea para interpretar datos y gestionar infodemias. Teorías relevantes incluyen:

Difusión de Innovaciones (Rogers): Cómo ideas/información son adoptadas por diferentes grupos (innovadores a rezagados).

Flujo en Dos Pasos/Multi-Pasos: La información fluye de fuentes a líderes de opinión y luego a redes más amplias (explica el rol de influencers).

Modelos de Umbral: Las personas adoptan algo cuando suficientes contactos lo hacen (explica la aceleración de la difusión).

Contagio Social / Modelos Epidémicos (SIR): Usan analogías de enfermedades para modelar la propagación de información.

Estas teorías explican la rápida difusión de (des)información y la importancia de la estructura de red y los factores sociales/psicológicos (Fallatah & Adekola, 2024; Tarkoma et al., 2020).

Principios de Ciencia de Datos Aplicados

Son esenciales para extraer conocimiento de datos digitales masivos y complejos.

Incluyen:

Minería de Datos: Descubrir patrones en grandes volúmenes de datos.

Procesamiento de Lenguaje Natural (PLN): Permite a las computadoras entender texto/habla (clave para redes sociales, noticias). Incluye clasificación, extracción de entidades (NER), análisis de sentimiento, etc.

Aprendizaje Automático (AA / ML): Algoritmos que aprenden de datos (supervisado, no supervisado). Fundamental para predicción y clasificación.

Aprendizaje Profundo (DL): Subconjunto de AA con redes neuronales profundas, potente para datos complejos (imágenes, texto).

La evaluación rigurosa de modelos es crucial. Un desafío es la interpretabilidad ("caja negra") de modelos complejos; la IA Explicable (XAI) busca abordarlo (Fallatah & Adekola, 2024; Tarkoma et al., 2020)..

Informática en Salud Pública

Aplica la informática (gestión de datos/información con tecnología) para mejorar la salud poblacional. Proporciona herramientas y marcos para la vigilancia.

Conceptos clave: vigilancia sindrómica, inteligencia epidémica (uso de fuentes diversas y herramientas informáticas para detectar amenazas), Infodemiología, Sistemas de Información Geográfica (SIG).

Principios fundamentales: Interoperabilidad (capacidad de sistemas para intercambiar y usar datos) y estándares de datos (esenciales para integrar fuentes digitales y tradicionales) e infraestructura adecuada.

Sinergia: La informática en salud pública actúa como el puente necesario para gestionar los datos digitales ruidosos (provenientes de la epidemiología digital), aplicar análisis avanzados (de la ciencia de datos) e integrar los hallazgos en sistemas y prácticas de salud pública efectivas, superando los silos de información (Dhewantara et al., 2019; Fallatah & Adekola, 2024; Shen et al., 2020; Tarkoma et al., 2020)..

La vigilancia digital se apoya en la epidemiología para el marco conceptual, las teorías de difusión para entender los datos sociales, la ciencia de datos para las herramientas de análisis, y la informática en salud pública para la gestión, integración y aplicación práctica de la información.

Marco Conceptual

Para trabajar en vigilancia digital, es crucial definir claramente los términos clave:

Detección Temprana y Brote

Detección Temprana: Es identificar un brote de enfermedad infecciosa lo más pronto posible para permitir una respuesta rápida y eficaz. Se centra en detectar la presencia real de la enfermedad, a diferencia de la "alerta temprana" que se enfoca en el riesgo de que aparezca (Aiello et al., 2020; Barros et al., 2020; Mavragani, 2020; Tarkoma et al., 2020).

Brote: Es la aparición de más casos de una enfermedad de los esperados normalmente en un lugar y tiempo específicos. Incluso un solo caso puede ser un brote si la enfermedad es nueva en la zona. Requiere investigación para confirmarse (Aiello et al., 2020; Barros et al., 2020; Mavragani, 2020; Tarkoma et al., 2020).

Datos y Análisis

Datos de Redes Sociales: Información generada por usuarios en plataformas como Twitter o Facebook (textos, imágenes, metadatos). Son voluminosos, rápidos y variados, pero pueden tener ruido y sesgos (Aiello et al., 2020; Barros et al., 2020; Mavragani, 2020; Tarkoma et al., 2020).

Fuentes Digitales: Un término más amplio que incluye redes sociales, pero también datos de búsquedas web (Google Trends), tráfico web, móviles, apps de salud, noticias online, etc. Su característica común es que no fueron creados originalmente para la vigilancia sanitaria (Aiello et al., 2020; Barros et al., 2020; Mavragani, 2020; Tarkoma et al., 2020).

Análisis de Datos: Proceso de examinar y modelar datos para obtener información útil. En vigilancia digital, usa técnicas como estadísticas, Procesamiento de Lenguaje Natural (PLN)

y Aprendizaje Automático (AA) (Aiello et al., 2020; Barros et al., 2020; Mavragani, 2020; Tarkoma et al., 2020).

Campos Relacionados

Infodemiología: Ciencia que estudia cómo la información (y desinformación) se distribuye en internet y afecta la salud pública y los comportamientos. Infoveillance es su aplicación específica a la vigilancia (Aiello et al., 2020; Barros et al., 2020; Mavragani, 2020; Tarkoma et al., 2020).

Vigilancia Digital: Uso de datos generados digitalmente (a menudo no para este fin) para vigilar enfermedades (monitorear, estimar presente/futuro). Puede complementar la vigilancia tradicional (Aiello et al., 2020; Barros et al., 2020; Mavragani, 2020; Tarkoma et al., 2020).

Epidemiología Digital: Campo más amplio que usa datos digitales (a menudo sin rigor estadístico inicial) para toda la investigación epidemiológica (no solo vigilancia). Enfatiza los retos metodológicos de usar estos datos (Aiello et al., 2020; Barros et al., 2020; Mavragani, 2020; Tarkoma et al., 2020).

Jerarquía: La Epidemiología Digital es el campo general. La Vigilancia Digital es una aplicación dentro de ella. La Infodemiología estudia la información digital, y la Infoveillance aplica eso a la vigilancia, solapándose con la Vigilancia Digital. Entender estas distinciones es clave. El enfoque principal del texto original es la Vigilancia Digital para la detección temprana de brotes (Aiello et al., 2020; Barros et al., 2020; Mavragani, 2020; Tarkoma et al., 2020).

Marco Normativo y Consideraciones Éticas

El uso de datos digitales, especialmente personales o generados por usuarios (como en redes sociales), para vigilar la salud pública presenta complejos desafíos éticos y legales. Es fundamental entender las directrices internacionales, las regulaciones de privacidad y las técnicas de protección de datos (Bruneau, 2020; De La Paz Lascano, 2022; Escarcena, 2021; Hunter et al., 2018).

Directrices Éticas de la OMS (2017)

La Organización Mundial de la Salud estableció 17 directrices éticas para la vigilancia en salud pública (Bruneau, 2020; De La Paz Lascano, 2022; Escarcena, 2021; Hunter et al., 2018)..

Obligaciones Clave: Los países deben tener sistemas de vigilancia adecuados con fines legítimos de salud pública, asegurar la calidad y seguridad de los datos, ser transparentes, considerar los valores comunitarios, minimizar riesgos y proteger a grupos vulnerables.

Compartir Datos: Existe la obligación de compartir datos con otras agencias (con salvaguardas) y es imperativo hacerlo en emergencias.

Consentimiento: Una directriz clave (y controvertida) establece que, bajo ciertas condiciones (datos fiables necesarios, protecciones adecuadas), los individuos tienen la obligación de contribuir a la vigilancia y no se requiere éticamente el consentimiento informado.

Limitación: Los datos identificables no deben usarse para fines punitivos o no relacionados con la salud.

Desafío Digital: Aplicar estas directrices a datos digitales (masivos, rápidos, a veces públicos pero personales, fáciles de vincular) es complejo. Aumentan los riesgos (brechas de seguridad, reidentificación, sesgos algorítmicos) y las expectativas de privacidad en línea son

variables. Se requiere una reflexión ética continua y adaptada a la tecnología (Bruneau, 2020; De La Paz Lascano, 2022; Escarcena, 2021; Hunter et al., 2018)..

Regulaciones de Privacidad (Ej. HIPAA en EE.UU.)

Leyes como HIPAA protegen la Información de Salud Protegida (PHI).

Regla de Privacidad HIPAA: Limita el uso y divulgación de PHI sin autorización del individuo, pero permite excepciones para tratamiento, pago, operaciones y fines de interés público, incluyendo actividades de salud pública como la vigilancia.

Principio del Mínimo Necesario: Usar solo la PHI indispensable.

Desidentificación HIPAA: La información deja de ser PHI si se desidentifica mediante Safe Harbor (eliminando 18 identificadores) o Determinación Experta (un experto evalúa el riesgo de reidentificación como muy pequeño).

Desafío Digital: HIPAA se diseñó para el sistema de salud formal, no para datos de fuentes públicas como redes sociales. Obtener autorización o aplicar el "mínimo necesario" a estos datos masivos es difícil. Los métodos de desidentificación pueden ser insuficientes o eliminar demasiada información útil. Esto crea una "zona gris" regulatoria que necesita guías más claras para el uso de datos digitales públicos (Bruneau, 2020; De La Paz Lascano, 2022; Escarcena, 2021; Hunter et al., 2018).

Técnicas de Anonimización y Desidentificación

Distinción: La desidentificación (estilo HIPAA) busca cumplir una norma eliminando identificadores (puede permitir reidentificación controlada), mientras que la anonimización busca hacer la reidentificación imposible o inviable (objetivo más estricto) (Bruneau, 2020; De La Paz Lascano, 2022; Escarcena, 2021; Hunter et al., 2018)..

Técnicas Comunes: Enmascaramiento, pseudonimización (tokens), generalización (rangos de edad, regiones), perturbación (ruido), datos sintéticos, K-anonimato (cada registro indistinguible de k-1 otros), y Privacidad Diferencial (marco matemático que añade ruido a los resultados del análisis para garantizar la privacidad individual, considerado el estándar de oro actual).

Consentimiento, Confianza y Participación Comunitaria

Consentimiento: Aunque la vigilancia de salud pública a menudo opera bajo una excepción al consentimiento informado (justificada por el bien común y salvaguardas), aplicarla a datos digitales públicos es debatido (Bruneau, 2020; De La Paz Lascano, 2022; Escarcena, 2021; Hunter et al., 2018)..

Transparencia: Es una obligación ética crucial informar al público sobre qué datos se usan, cómo y por qué, para construir confianza.

Confianza Pública: Es esencial pero frágil. Se daña por falta de transparencia, brechas de seguridad o usos inapropiados.

Participación Comunitaria: Involucrar a las comunidades asegura la relevancia y aceptación de la vigilancia.

Interconexión: La confianza, basada en la transparencia y el respeto ético, es fundamental para la legitimidad y sostenibilidad de la vigilancia digital. Abordar la ética no es solo cumplir normas, sino una necesidad operativa.

Metodología

Protocolo

Este estudio se realizó siguiendo las directrices del Joanna Briggs Institute para revisiones de alcance y el marco PRISMA-SCR (Munn et al., 2020; Tricco et al., 2018). Se desarrolló un protocolo que detalla los criterios de elegibilidad, estrategias de búsqueda, procedimientos de selección de estudios, extracción de datos y análisis descriptivo, con el fin de mapear y sintetizar la literatura sobre vigilancia digital de enfermedades.

Criterios de Elegibilidad

Se incluyeron estudios empíricos (observacionales, experimentales y de modelado) y revisiones sistemáticas o de alcance que:

Reporten métodos y resultados de vigilancia digital de enfermedades infecciosas (por ejemplo, influenza, COVID-19, RSV, dengue, etc.).

Describan características del diseño del estudio, fuentes digitales, métodos de detección, ventajas temporales, precisión y correlación con sistemas tradicionales de vigilancia.

Se hayan publicado en revistas revisadas por pares y estén disponibles en texto completo, sin restricciones de idioma.

Se excluyeron estudios que:

No aporten datos empíricos o que se centren exclusivamente en modelos teóricos sin validación.

Se enfoquen únicamente en vigilancia tradicional sin integrar datos digitales.

Presenten resultados poco claros o insuficientes para extraer la información requerida.

Pregunta PCC (Población, Concepto, Contexto)

Población (P): Estudios que involucren el uso de datos digitales (ej., redes sociales, motores de búsqueda, aplicaciones móviles) para la vigilancia de enfermedades en poblaciones humanas.

Concepto (C): Características, desempeño y métodos de vigilancia digital de enfermedades; es decir, se extraerán variables que describan el diseño del estudio, las fuentes de datos, el método de detección, la ventaja temporal, la tasa de detección, la precisión, entre otros.

Contexto (C): Estudios publicados en revistas científicas revisadas por pares que aborden vigilancia digital de brotes o epidemias a nivel local, regional, nacional o global.

Estrategia de Búsqueda

Se diseñó una estrategia de búsqueda integral para identificar estudios relevantes en bases de datos clave (PubMed, Scopus, Web of Science, Springer, SciELO, Science Direct, Google Scholar, and Redalyc.). La búsqueda se centró en términos clave relacionados con:

("digital surveillance" OR "infodemiology") AND ("infectious diseases" OR "epidemic") AND ("social media" OR "Google Trends" OR "Twitter") AND ("machine learning" OR "forecasting")

Selección de Estudios y Extracción de Datos

Dos revisores independientes (Yamil Liscano y Luis Anillo) examinaron los títulos y resúmenes de los estudios identificados para determinar su elegibilidad. Los artículos preseleccionados se sometieron a una evaluación en texto completo para confirmar el cumplimiento de los criterios de inclusión y exclusión. En caso de discrepancias, se resolvieron mediante consenso o consulta a un tercer revisor.

Se diseñó un formulario estructurado para la extracción de datos, en el que se recopilaron las siguientes variables (con sus correspondientes definiciones adaptadas para vigilancia digital):

- Autor y año
- Diseño del estudio
- Fuentes de datos digitales
- Método de comparación empleado
- Ámbito geográfico del estudio
- Técnicas o algoritmos utilizados para la detección de la señal digital
- Ventaja temporal en la detección
- Indicadores o tasas de detección reportadas
- Medidas de precisión y desempeño
- Periodo de recolección de datos
- Tipo de enfermedad o brote
- Plataformas y herramientas digitales específicas
- Métodos de preprocesamiento de datos
- Algoritmos o técnicas analíticas
- Métricas estadísticas y de desempeño
- Resolución espacial y granularidad temporal
- Integración con sistemas de vigilancia tradicionales
- Proceso de selección de palabras clave
- Medición del impacto mediático
- Características demográficas y de uso

La figura 1 se creó con el paquete online *R package* PRISMA2020 (Haddaway et al., 2022) (https://estech.shinyapps.io/prisma_flowdiagram/, accedido el 24 Febrero 2025).

Análisis Estadístico

Se realizará un análisis descriptivo univariante de todas las variables extraídas, utilizando software R (versión 4.3.0) (accedido el 15 de Marzo de 2025). Los resultados se presentarán en tablas y gráficos que ilustren las tendencias en el diseño de estudios, las fuentes de datos, los métodos de detección y otros aspectos relevantes usando la librería ggplot2.

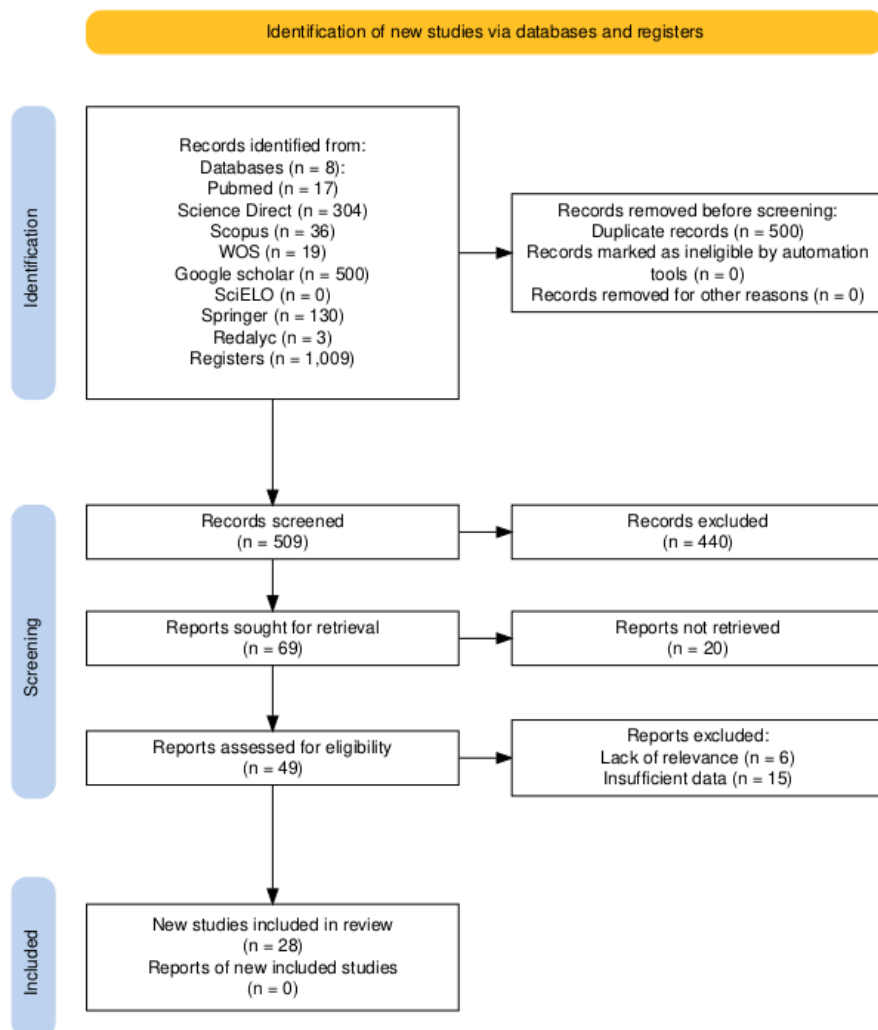
Resultados

Información General del Estudio

Un total de 1009 registros fueron identificados en las distintas fuentes de búsqueda (PubMed, Science Direct, Scopus, WOS, SciELO, Springer, Redalyc y *Google Scholar*). Tras eliminar 500 duplicados, quedaron 509 registros. De estos, 440 se excluyeron por título y resumen, de modo que 69 pasaron a la fase de *screening*. En esta etapa, se descartaron 20 registros, quedando 49 informes potencialmente relevantes. Finalmente, se excluyeron 6 informes por falta de relevancia, de modo que 28 estudios se incluyeron en la revisión (ver figura 1).

Figura 1

Diagrama PRISMA



Los estudios sobre vigilancia digital en salud pública destacan por una notable diversidad metodológica, reflejando una evolución hacia enfoques más sofisticados. Como se muestra en la Tabla 1, las investigaciones incluyen desde análisis empíricos cuantitativos comparativos (por ejemplo, Dai et al., 2020), estudios observacionales (Lamos et al., 2021), análisis retrospectivos y de cohortes (Timpka et al., 2014), hasta estudios exploratorios (van de Belt et al., 2018). Esta

variedad evidencia cómo las técnicas estadísticas y de modelado se han adaptado progresivamente a la creciente disponibilidad de datos digitales, incorporando métodos avanzados como aprendizaje por transferencia, análisis bayesiano y series temporales para mejorar la predicción y gestión de brotes.

Tabla 1

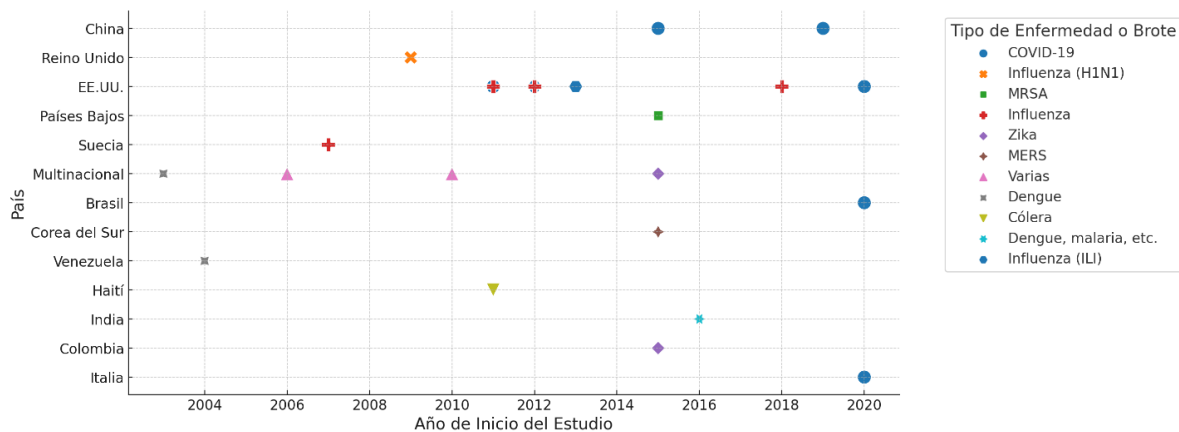
Características Generales de los Estudios Incluidos

Autor y año	Study Design	Periodo de recolección de datos	Locación	Tipo de enfermedad o brote
Dai et al. (2020)	Empírico cuantitativo comparativo	2015-2020	China	COVID-19
Lamos et al. (2010)	Empírico cuantitativo	2009	Reino Unido	Influenza (H1N1)
Lamos et al. (2021)	Observacional/modelado	2011-2020	EE.UU., Reino Unido, Australia, etc.	COVID-19
van de Belt et al. (2018)	Exploratorio comparativo	2015-2017	Países Bajos	MRSA
Timpka et al. (2014)	Cohorte abierta	2007-2012	Suecia	Influenza
McGough et al. (2017)	Retrospectivo pronóstico multivariable	2015-2016	Latinoamérica	Zika
Yousefinaghan i et al. (2021)	Observacional, retrospectivo y predictivo	2020	EE.UU. y Canadá	COVID-19
Wittwer et al. (2023)	Transversal comparativo	2020	Brasil	COVID-19

Shin et al. (2016)	Observacional correlación	2015	Corea del Sur	MERS
Yan et al. (2017)	Revisión sistemática	2006-2016	Internacional	Varias
Strauss et al. (2017)	Observacional correlacional	2004-2014	Venezuela	Dengue
Chunara et al. (2012)	Observacional	2011	Haití	Cólera
Barboza et al. (2014)	Evaluación cuantitativa	2010	Internacional	Varias
Kogan et al. (2021)	Alerta temprana	2020	EE.UU.	COVID-19
Verma et al. (2018)	Transversal correlativo	2016	India	Dengue, malaria, etc.
Santillana et al. (2015)	Aprendizaje automático	2013-2015	EE.UU.	Influenza (ILI)
Majumder et al. (2016)	Retrospectivo	2015-2016	Colombia	Zika
Samaras et al. (2020)	Comparativo	Temporada de influenza	Grecia	Influenza
Li et al. (2021)	Retrospectivo	2020	EE.UU.	COVID-19
Feldman et al. (2019)	Desarrollo de base de datos	10 meses	Global	114 enfermedades
Sharpe et al. (2016)	Comparativo retrospectivo	2012-2015	EE.UU.	Influenza (ILI)
Porcu et al. (2023)	Retrospectivo	2020-2021	Italia	COVID-19
Sun et al. (2018)	Observacional retrospectivo	2012-2016	Boston, EE.UU.	Influenza
Chan et al. (2011)	Monitoreo en tiempo real	2003-2010	Bolivia, Brasil, India, etc.	Dengue

Wang et al. (2023)	Predicción de brotes	5 años	Japón, Alemania, Bélgica	RSV
Alessa & Faezipour (2019)	Observacional retrospectivo	2018	EE.UU. (Connecticut)	Influenza
Broniatowski et al. (2013)	Infoveillance observacional	2012-2013	EE.UU. (Nacional y NYC)	Influenza
Shen et al. (2020)	Observacional retrospectivo	2019-2020	China	COVID-19

Respecto a la temporalidad de la recolección de datos, la Figura 2 ilustra que existe una considerable amplitud, desde estudios tempranos como el de Chan et al. (2011) sobre dengue (desde 2003), hasta investigaciones recientes vinculadas con COVID-19 (Lampos et al., 2021; Kogan et al., 2021; Shen et al., 2020). Esta diversidad temporal responde a las características particulares de cada brote. Estudios sobre influenza suelen analizar datos de múltiples temporadas, mientras que investigaciones sobre brotes repentinos como COVID-19 o Zika tienden a centrarse en periodos más acotados, reflejando la urgencia y dinámica emergente de estas epidemias.

Figura 2*Distribución Temporal y Geográfica de los Estudios*

La diversidad geográfica también es un aspecto clave destacado en la Tabla 1. Se observan estudios con alcance local, como el de Wittwer et al. (2023) en ciudades brasileñas, o el de Broniatowski et al. (2015) en un hospital de Baltimore. Otros trabajos poseen enfoques nacionales realizados en países como China, Reino Unido, Estados Unidos, Italia e India, y también hay investigaciones internacionales que abarcan múltiples países, como las de Yan et al. (2017) y Feldman et al. (2019). Esta amplitud geográfica permite adaptar las metodologías a contextos específicos, considerando aspectos como la conectividad a internet, la densidad poblacional o particularidades culturales en la comunicación digital.

En cuanto a las enfermedades estudiadas, predominan claramente las respiratorias y virales, especialmente COVID-19 e influenza, debido a su relevancia epidemiológica y disponibilidad de datos digitales. Sin embargo, también se han estudiado otras enfermedades como Zika, MRSA, MERS, dengue y cólera. Investigaciones amplias como la de Feldman et al. (2019), que cubre múltiples patologías, validan la aplicabilidad general de estas metodologías.

La Figura 1 muestra claramente tendencias cronológicas en los estudios realizados. Antes de 2010 predominaban influenza y dengue; entre 2010 y 2015 aumentaron los estudios sobre patógenos adicionales como MERS, cólera y MRSA. Posteriormente, entre 2015 y 2016, destacaron investigaciones sobre Zika en Latinoamérica y MERS en Asia. Finalmente, desde 2020 se observa un notable incremento de estudios relacionados con COVID-19 a nivel mundial, reflejando la importancia de esta enfermedad en la vigilancia epidemiológica digital.

Fuentes de Datos y Plataformas Digitales

Los estudios sobre vigilancia digital en salud pública reflejan un ecosistema amplio y diverso de herramientas digitales, destacando claramente el predominio de Google y Twitter, como se aprecia en la Figura 3 y la Tabla 2. Google (búsquedas y tendencias) es la fuente más utilizada, apareciendo con la mayor frecuencia (20 estudios). Esto coincide con múltiples investigaciones que utilizan Google Trends, Google Flu Trends o consultas específicas de búsqueda para detectar brotes tempranos (Dai et al., 2020; Lampos et al., 2021; Verma et al., 2018; Porcu et al., 2023) o correlacionar picos de búsqueda con datos oficiales (Strauss et al., 2017). Su uso también es comúnmente combinado con otras plataformas como Wikipedia y Twitter para fortalecer la vigilancia epidemiológica (Sharpe et al., 2016).

Tabla 2

Fuentes de Datos Digitales y Plataformas Empleadas

Autor y año	Plataformas y herramientas digitales específicas	Integración con sistemas de vigilancia tradicionales
Dai et al. (2020)	Motor de búsqueda Baidu	Comparación con el sistema tradicional de reporte de casos
Lampos et al. (2010)	Twitter	Calibración del “flu-score” con datos de HPA

Autor y año	Plataformas y herramientas digitales específicas	Integración con sistemas de vigilancia tradicionales
Lamos et al. (2021)	Google Search y datos de noticias	Comparación con datos oficiales de casos y muertes
van de Belt et al. (2018)	Coosto (monitoreo de redes sociales) y Google Trends	Comparación con notificaciones oficiales en el sistema SO-ZI/AMR
Timpka et al. (2014)	Google Flu Trends, Healthcare Direct/1177, Google Analytics	Comparación con datos clínicos y de laboratorio de influenza
McGough et al. (2017)	Google Search, Twitter, HealthMap	Integración con datos de Zika reportados por PAHO y ministerios de salud
Yousefinaghani et al. (2021)	Twitter API y Google Trends	Comparación con datos oficiales (Johns Hopkins COVID-19)
Wittwer et al. (2023)	Brazil Sem Corona y datos de GitHub	Integración con datos PS y TS para mejorar la predicción
Shin et al. (2016)	Google Trends, Topsy	Comparación con datos oficiales de MERS
Yan et al. (2017)	Google Flu Trends, Google Trends, Baidu, Twitter, ProMED-mail, HealthMap	Discusión sobre complementariedad con sistemas tradicionales
Strauss et al. (2017)	Google Dengue Trends	Comparación y propuesta de complementariedad con el sistema de vigilancia
Chunara et al. (2012)	HealthMap y Twitter	Comparación con datos oficiales del MSPP
Barboza et al. (2014)	Argus, BioCaster, GPHIN, HealthMap, MedISys, ProMED-mail	Evaluación comparativa con datos oficiales del BHI
Kogan et al. (2021)	Google Trends, Twitter, UpToDate, GLEAM, Apple Mobility, Cuebiq, Kinsa Thermometer	Integración de proxies digitales con casos, muertes e ILI

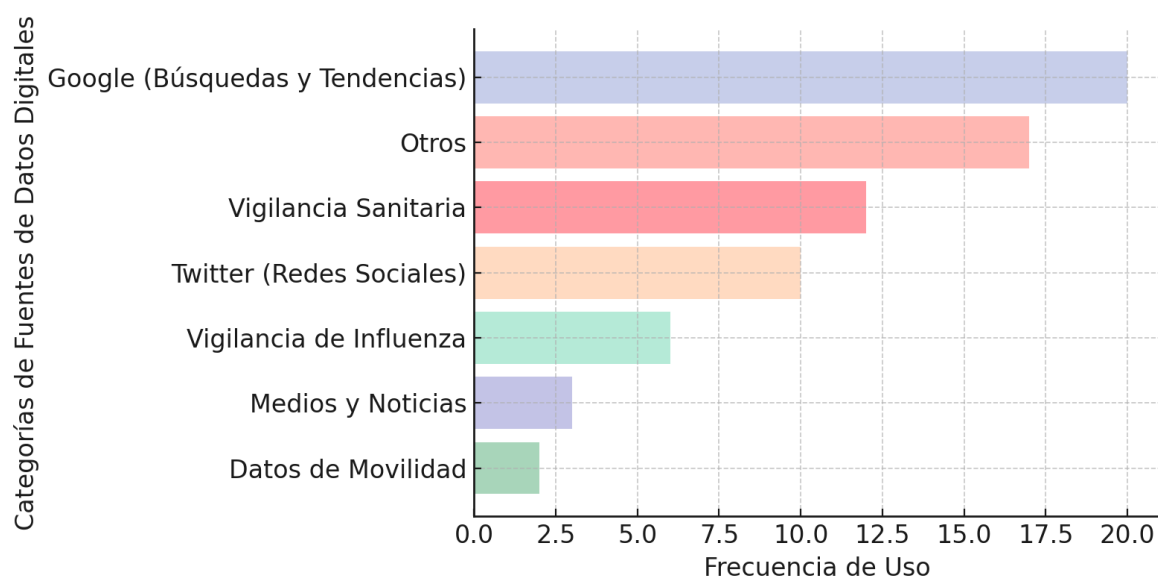
Autor y año	Plataformas y herramientas digitales específicas	Integración con sistemas de vigilancia tradicionales
Verma et al. (2018)	Google Trends y Google Correlate	Comparación con el sistema de vigilancia del IDSP
Santillana et al. (2015)	Google Trends, Twitter, athenahealth, FluNearYou	Contraste de predicciones con reportes del CDC
Majumder et al. (2016)	HealthMap y Google Trends	Validación con datos oficiales del INS
Samaras et al. (2020)	Google Trends, Twitter API (Tweepy y Pytrends)	Comparación con datos oficiales de influenza en Europa
Li et al. (2021)	Twitter Standard Search API	Comparación con sistemas oficiales basados en búsquedas y noticias
Feldman et al. (2019)	GDELT Global Knowledge Graph y Google Translate API	Comparación con reportes de la WHO (DON)
Sharpe et al. (2016)	Google Flu Trends, HealthTweets, Wikipedia	Comparación con reportes oficiales del CDC
Porcu et al. (2023)	Google Trends	Validación con datos de RT-PCR
Sun et al. (2018)	Google Trends, Twitter, athenahealth, Flu Near You	Validación con datos del Boston Public Health Commission
Chan et al. (2011)	Google Search queries	Comparación con datos de ministerios de salud y WHO
Wang et al. (2023)	Google Trends	Complemento para la vigilancia clínica
Alessa & Faezipour (2019)	Twitter	Validación con datos del CDC y hospitales
Broniatowski et al. (2013)	Twitter API (HealthTweets y Google Flu Trends)	Validación con reportes del CDC y Departamento de Salud de NYC
Shen et al. (2020)	Weibo	Comparación con datos oficiales del China CDC

Autor y año	Plataformas y herramientas digitales específicas	Integración con sistemas de vigilancia tradicionales
Broniatowski et al (2015)	Twitter (HealthTweets)	Comparación con datos hospitalarios (casos de laboratorio e ILI en ED)

Twitter o ahora llamado “X” y otras redes sociales ocupan una posición complementaria significativa, siendo la segunda fuente más utilizada según la Figura 3, con alrededor de 12 menciones. Diversos estudios aprovechan la plataforma para analizar tendencias mediante el conteo de términos relacionados con síntomas o enfermedades específicas (Lamos et al., 2010; McGough et al., 2017; Yousefinaghani et al., 2021; Samaras et al., 2020). Además, Twitter se contrasta frecuentemente con información oficial proporcionada por instituciones como el CDC, destacando su utilidad para complementar datos tradicionales mediante herramientas avanzadas de filtrado y geolocalización (Santillana et al., 2015; Alessa & Faezipour, 2019; Shin et al., 2016).

Figura 3

Frecuencia de Uso de Diferentes Fuentes de Datos



Las plataformas especializadas en vigilancia sanitaria, como HealthMap, ProMED-mail, GPHIN o Argus, también tienen una relevancia notable, confirmada en la Tabla 2. Estas plataformas son empleadas frecuentemente para obtener alertas rápidas a partir de agregadores de noticias o reportes institucionales (Barboza et al., 2014; Chunara et al., 2012; Majumder et al., 2016). La mayoría de los estudios integra estas alertas con sistemas oficiales de vigilancia sanitaria, como reportes del ministerio de salud o de la OMS, fortaleciendo la validez y precisión de las observaciones (McGough et al., 2017; Chunara et al., 2012).

Otras fuentes digitales menos utilizadas, pero emergentes, incluyen medios y bases de noticias, como GDELT o BioCaster, que permiten detectar brotes emergentes mediante análisis automatizados de noticias digitales (Lamos et al., 2021; Feldman et al., 2019; Barboza et al., 2014). Además, se ha incrementado el interés en incorporar nuevas fuentes, como datos de movilidad (Apple Mobility, Cuebiq) y termómetros inteligentes (plataforma Kinsa Thermometer), especialmente en estudios sobre COVID-19 para generar alertas tempranas (Kogan et al., 2021).

Un factor común en casi todos estos estudios, es la comparación sistemática y validación con sistemas oficiales. Este patrón evidencia que las fuentes digitales funcionan predominantemente como complemento y rara vez como reemplazo de la vigilancia epidemiológica tradicional. Los estudios sobre influenza y COVID-19 subrayan esta complementariedad, mostrando que la vigilancia digital puede anticipar tendencias epidemiológicas, pero que la validación final depende del cotejo con registros clínicos y reportes epidemiológicos oficiales (Timpka et al., 2014; Santillana et al., 2015; Dai et al., 2020; Yousefinaghani et al., 2021).

Métodos y Técnicas Analíticas

Los estudios sobre vigilancia digital en salud pública presentan una estructura metodológica común claramente identificable en la Figura 4 y detallada en la Tabla 3. Aunque existen variaciones según la enfermedad estudiada o la fuente de datos empleada, es posible identificar etapas compartidas que garantizan una detección eficaz y temprana de brotes.

Tabla 3

Métodos de Análisis y Procesamiento de Datos

Autor y año	Método de comparación	Método de detección	Preprocesamiento	Técnicas analíticas
Dai et al. (2020)	Análisis de correlación entre picos anómalos y reportes oficiales	Incremento anormal en ILI y búsquedas ('pneumonia', 'SARS')	Suavizado (media móvil 7 días)	ANOVA, regresión lineal, correlación
Lamos et al. (2010)	Comparación de 'flu-score' en tweets vs tasas de ILI	Cálculo de 'flu-score' a partir de tweets	Eliminación de stop words, stemming, suavizado	Regresión lineal, LASSO, aprendizaje supervisado
Lamos et al. (2021)	Comparación de consultas online con datos oficiales de COVID-19	Modelos unsupervised y transfer learning con síntomas	Normalización y ponderación de síntomas	Elastic net, Gaussian processes, correlación
van de Belt et al. (2018)	Comparación de brotes detectados en redes con reportes oficiales	Detección en redes sociales y Google Trends	Umbrales en redes y Google Trends	Estadística descriptiva, análisis ROC, correlación
Timpka et al. (2014)	Comparación de datos eHealth con casos	Correlación de datos eHealth con datos clínicos	Ajuste de efectos	Regresión lineal, autoregresivos, correlación

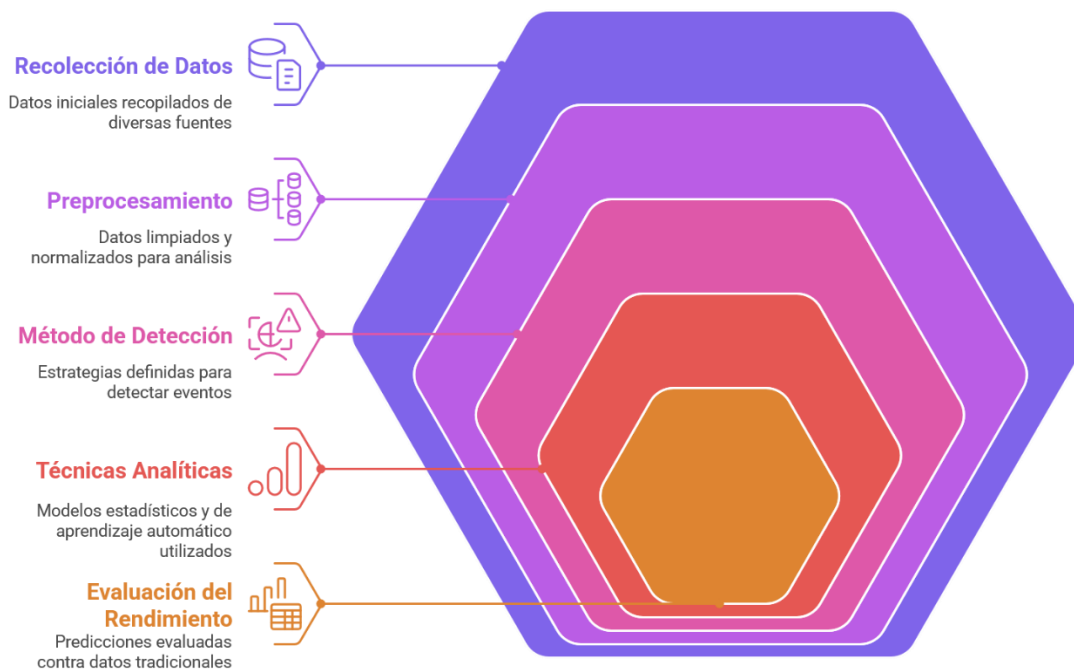
Autor y año	Método de comparación	Método de detección	Preprocesamiento	Técnicas analíticas
	clínicos y de laboratorio		semanales, detrending	
McGough et al. (2017)	Modelos predictivos de casos de Zika con datos digitales	Predicción de casos mediante señales digitales	Transformaciones log y normalización	Elastic net, validación cruzada, autoregresivos
Yousefin aghani et al. (2021)	Comparación de series temporales digitales con casos de COVID-19	Análisis de anomalías en tweets y búsquedas	Filtrado por keywords y geolocalización	Análisis de anomalías, regresión, validación
Wittwer et al. (2023)	Comparación de tasas de infección autoinformadas vs oficiales	Estimación de tasas de infección en autoinformes	LOESS para suavizar fluctuaciones	Modelos autoregresivos, AIC, combinación de variables
Shin et al. (2016)	Correlación entre datos digitales y casos oficiales	Correlación de lag entre datos digitales y casos	Normalización y selección de palabras	Spearman y análisis de retardo
Yan et al. (2017)	Análisis de correlación y detección de señales digitales	Detección de señales digitales en reportes oficiales	Categorización y eliminación de ruido	Correlación, algoritmos bayesianos, detección de señales
Strauss et al. (2017)	Comparación de vigilancia digital con casos reportados de dengue	Vigilancia digital basada en búsquedas de dengue	Normalización y conversión de volúmenes	Regresión lineal, análisis de correlación
Chunara et al. (2012)	Análisis de correlación entre tweets y reportes de cólera	Análisis de reportes en HealthMap y Twitter	Filtrado y selección de términos clave	Ajuste exponencial,

Autor y año	Método de comparación	Método de detección	Preprocesamiento	Técnicas analíticas
				ecuación de Lotka-Euler
Barboza et al. (2014)	Evaluación de biosupervisión con señales en medios	Búsquedas en medios con validación humana	Filtrado manual y eliminación de duplicados	Regresión Poisson, cálculos de tasas
Kogan et al. (2021)	Comparación de proxies digitales con datos de casos y muertes	Modelado de proxies digitales y datos oficiales	Suavizado y escalado de proxies digitales	Crecimiento exponencial, media armónica, correlación
Verma et al. (2018)	Correlación entre patrones de búsqueda y brotes en India	Identificación de términos en Google Correlate	Selección de términos en Google Correlate	Análisis de correlación y series temporales
Santillana et al. (2015)	Predicción de ILI combinando múltiples fuentes digitales	Predicción de actividad ILI con múltiples proxies	Normalización y mapeo de fuentes digitales	Regresión LASSO, SVM, AdaBoost
Majumder et al. (2016)	Estimación de transmisión de Zika con datos digitales	Modelado de transmisión de Zika con IDEA	Escalado y suavizado de Google Trends	Optimización no lineal, minimización de SSD
Samaras et al. (2020)	Modelado predictivo de influenza con ARIMA	Predicción de actividad epidémica con ARIMA	Eliminación de duplicados en Twitter	Modelos ARIMA(X), análisis predictivo
Li et al. (2021)	Clasificación de tweets de COVID-19 y análisis de lead time	Clasificación de tweets como alerta de COVID-19	Tokenización y lematización de tweets	Clasificación supervisada y análisis de sentimiento

Autor y año	Método de comparación	Método de detección	Preprocesamiento	Técnicas analíticas
Feldman et al. (2019)	Validación de detección de brotes con reportes de WHO	Detección de brotes en artículos de noticias	Traducción automática y filtrado por etiquetas	Naïve Bayes, SVM, LSTM bidireccional
Sharpe et al. (2016)	Detección de cambios en series con métodos bayesianos	Identificación de puntos de cambio en series	Normalización y agrupación semanal	Modelos bayesianos de cambio de punto
Porcu et al. (2023)	Detección de outliers en búsquedas con ARMA y EWMA	Detección de señales epidémicas en búsquedas	Ajuste de escalas de 0 a 100	ARMA, EWMA, detección de outliers
Sun et al. (2018)	Comparación de modelos ARGO vs autoregresivos simples	Predicción de influenza con modelos ensemble	Filtrado de términos irrelevantes	Regresión multivariable y método ensemble
Chan et al. (2011)	Ajuste de modelos lineales con búsquedas de dengue	Regresión lineal con búsquedas de dengue	Reemplazo de picos espurios	Regresión lineal univariada
Wang et al. (2023)	Correlación entre Google Trends y vigilancia clínica	Definición de umbrales con Moving Epidemic Method	Exclusión de años atípicos (2020-2021)	Moving Epidemic Method (MEM)
Alessa & Faezipour (2019)	Clasificación de tweets con FastText y regresión lineal	Regresión y clasificación de tweets	Stemming y eliminación de stopwords	FastText y regresión lineal
Broniato wski et al. (2013)	Filtrado de tweets para detección de influenza	Clasificación supervisada de tweets de influenza	Filtrado en etapas de tweets	SVM, regresión logística

Autor y año	Método de comparación	Método de detección	Preprocesamiento	Técnicas analíticas
Shen et al. (2020)	Causalidad de Granger entre 'sick posts' y conteos de casos	Predicción de casos con Granger y modelos supervisados	Clasificación en 'sick' vs otros	Clasificador random forest y regresión OLS
Broniato wski et al (2015)	Estimación de prevalencia de influenza con tweets y conteos	Estimación de prevalencia de ILI con Twitter	Normalización de volúmenes de tweets	Análisis ARIMAX y regresión logística

La primera etapa corresponde a la recolección de datos, donde destaca una diversidad significativa de fuentes digitales (ver figura 4). Estas incluyen motores de búsqueda, redes sociales, plataformas especializadas de vigilancia epidemiológica y datos de movilidad, entre otras. El objetivo central es obtener señales tempranas que complementen los sistemas tradicionales de vigilancia epidemiológica.

Figura 4*Diagrama de Flujo Metodológico*

Posteriormente, en la fase de preprocesamiento, es habitual aplicar técnicas de limpieza y normalización de datos, tales como escalas estándar (0–100, min-max, z-score) y suavizado mediante medias móviles. Además, algunos estudios realizan clasificación previa de la información para seleccionar contenido relevante frente al ruido digital, como ocurre en los trabajos de Li et al. (2021) y Feldman et al. (2019). Este proceso garantiza una calidad óptima de los datos para las fases analíticas posteriores.

La tercera etapa, centrada en la detección, utiliza métodos variados como análisis de correlaciones con datos oficiales, análisis de desfases temporales (lag correlations) y causalidad

de Granger. Estas técnicas permiten identificar patrones predictivos y desfases temporales respecto a los reportes oficiales (Lampos et al., 2021; Yousefinaghani et al., 2021; Shin et al., 2016). También se emplean técnicas avanzadas como modelos predictivos basados en causalidad o métodos estadísticos para mejorar la anticipación de brotes.

La cuarta fase, técnicas analíticas, implica una variedad de métodos estadísticos y de aprendizaje automático. Entre estos destacan regresiones lineales y modelos ARIMA (Samaras et al., 2020; Porcu et al., 2023), así como enfoques más sofisticados que incluyen regresiones supervisadas, modelos ensamblados y enfoques bayesianos (Sharpe et al., 2016; Kogan et al., 2021; Santillana et al., 2015). Estos métodos permiten realizar análisis robustos y precisos de las señales epidemiológicas capturadas.

El gráfico 5 muestra una matriz donde cada fila representa un estudio ordenado por año y cada columna una técnica analítica utilizada. Se observa una alta recurrencia de métodos como la regresión lineal y técnicas correlacionales, lo que indica su amplia aceptación en el campo de la vigilancia digital. También se identifican combinaciones como la integración de *machine learning* con métodos estadísticos clásicos, reflejando avances metodológicos en el abordaje de problemas complejos. Este gráfico es especialmente útil para revisiones metodológicas, ya que evidencia tendencias y brechas en el uso de técnicas analíticas, orientando futuras investigaciones. Las métricas más frecuentes incluyen coeficientes de correlación (Pearson, Spearman), errores (RMSE, MAE, MAPE) y medidas de clasificación como precisión, *recall* y F1-score (Li et al., 2021; Feldman et al., 2019). Además, la validación con datos epidemiológicos oficiales es casi universal, lo que resalta la complementariedad entre enfoques digitales y vigilancia tradicional.

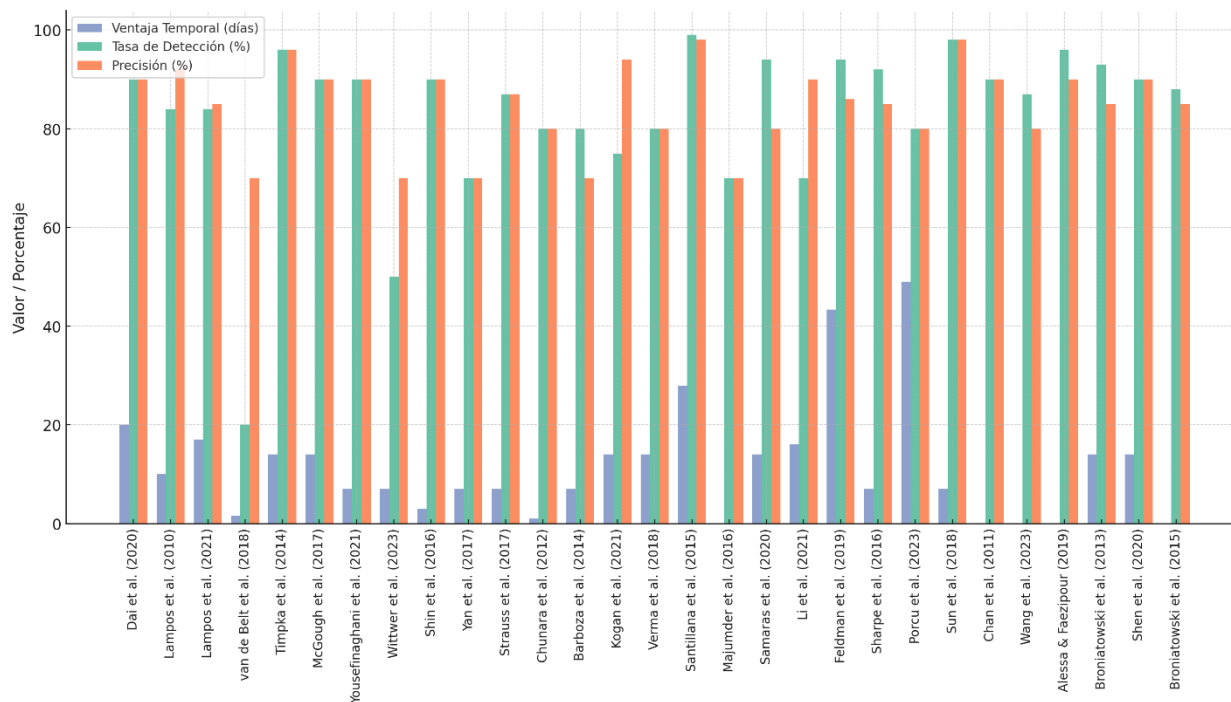
Tabla 4*Desempeño en Detección Temprana y Precisión*

Autor y año	Ventaja Temporal	Tasa de Detección	Precisión
Dai et al. (2020)	20 días antes de la alerta oficial	Altas correlaciones, sin tasa específica	Altos coeficientes de correlación
Lampos et al. (2010)	Tweets en horas, HPA tarda 1-2 semanas	Correlaciones 81.78%-85.56%	Validación cruzada ~89-94%
Lampos et al. (2021)	Casos: 16.7 días antes, muertes: 22.1 días antes	Correlación $r \approx 0.82-0.85$	Evaluada con AUC y MAE
van de Belt et al. (2018)	Brotos detectados 1-2 días antes	Sensibilidad 20%, especificidad 96%	AUC, sensibilidad, especificidad
Timpka et al. (2014)	GFT 2 semanas antes, telenursing varía	GFT $r=0.96$, telenursing $r \approx 0.95-0.97$	Pearson r, RMSE
McGough et al. (2017)	Pronósticos de 1-3 semanas antes	Medida por error predictivo (rRMSE)	RMSE, rRMSE, Pearson ρ
Yousefinaghani et al. (2021)	83% de olas detectadas 1 semana antes	Síntomas detectados 100% en EE.UU. Cat. I	RMSE, MAE, correlaciones $>75\%$
Wittwer et al. (2023)	Ventaja depende de la participación	Correlación alta en ciudades con buena participación	RMSE, MAE, Pearson correlation
Shin et al. (2016)	3-4 días antes de la confirmación	Correlaciones >0.7 , hasta 0.9	Correlaciones significativas ($p < 0.05$)
Yan et al. (2017)	1-12 días antes de reportes oficiales	Alerta 1-12 días antes, correlación variable	Moderada a alta según método
Strauss et al. (2017)	Alerta temprana antes de actualización	$r=0.87$ en semanas epidémicas	$R^2=0.75$ en regresión

Autor y año	Ventaja Temporal	Tasa de Detección	Precisión
Chunara et al. (2012)	Actualización diaria, oficial se retrasa 1-2 días	$\rho \approx 0.80$ en fases de crecimiento	Variabilidad en Re (1.54 a 6.89)
Barboza et al. (2014)	Detecta eventos antes de la publicación	C-DR 83-95%, I-DR 47-92%	Diferencias estadísticas en I-Se
Kogan et al. (2021)	Incrementos en casos 2-3 semanas antes	Sensibilidad combinada hasta 0.75	Precisión 0.90-0.98 en proxies
Verma et al. (2018)	Google Trends anticipa 2-3 semanas	$r > 0.80$ para chikungunya y dengue	Chikungunya $r = 0.82-0.87$
Santillana et al. (2015)	Predicción hasta 4 semanas antes	Predicción real-time $r = 0.989$	RMSE 0.176% ILI, MAPE reducido
Majumder et al. (2016)	Estimaciones en tiempo casi real	No se reporta tasa, estimación de R_0	Buen ajuste en modelos SSD
Samaras et al. (2020)	Búsquedas y tweets anticipan 2-3 semanas	Pearson $R \approx 0.933-0.943$	MAPE $\approx 18.7-22.6\%$
Li et al. (2021)	Detecta señales 16 días antes	Estrategia signal identifica alertas	Alta precisión en clasificación
Feldman et al. (2019)	Brotos detectados 43.4 días antes en promedio	94% brotos detectados antes de WHO	Recall 88.8%, Precisión 86.1%
Sharpe et al. (2016)	Google alerta cambios 1-2 semanas antes	Google: Sens. 92%, PPV 85%	Google con mejor desempeño
Porcu et al. (2023)	Epidemias detectadas 7-8 semanas antes	PPV 80% en Lombardy, $< 50\%$ en Marche	Alta correlación en conectividad alta
Sun et al. (2018)	Nowcasting y forecast a 1 semana	Correlaciones 0.98 (nowcast) y 0.94 (forecast)	RMSE, MAE, MAPE bajos
Chan et al. (2011)	Datos disponibles en tiempo real	Correlaciones 0.82-0.99	Buen ajuste en correlaciones

Autor y año	Ventaja Temporal	Tasa de Detección	Precisión
Wang et al. (2023)	Datos casi inmediatos	Japón $r=0.87$, Alemania $r=0.65$	Buena estimación de umbrales
Alessa & Faezipour (2019)	Casi en tiempo real	Correlación 96.29% con CDC	F-measure 89.9%
Broniatowski et al. (2013)	Tweets disponibles hasta 2 semanas antes	Nacional $r=0.93$, municipal $r=0.88$	Menor MAE en modelo infección
Shen et al. (2020)	Predice casos 14 días antes	Sick posts explican 12.8% varianza	Coefficientes estandarizados altos
Broniatowski et al. (2015)	Tweets antes que datos oficiales	Correlación alta a nivel municipal	85% exactitud en tendencia

La ventaja temporal (lead time) proporcionada por estas herramientas digitales es notable, con detecciones anticipadas que varían desde pocos días hasta varias semanas en comparación con los sistemas tradicionales (ver figura 6). Aunque la mayoría de los estudios reporta adelantos de pocos días hasta varias semanas, algunos casos destacan significativamente, como Feldman et al. (2019), que documenta ventajas de hasta 43 días. Estas diferencias se explican principalmente por la diversidad de fuentes utilizadas (por ejemplo, búsquedas en Google frente a redes sociales) y la metodología empleada (detección de picos, análisis correlacionales o algoritmos supervisados).

Figura 6*Comparación del Lead Time y Precisión Entre Estudios*

En cuanto a la tasa de detección, existe una considerable variabilidad. Mientras algunos estudios presentan tasas altas de detección, como las investigaciones que utilizan Google Trends y Twitter, otros reportan resultados más moderados o inconsistentes, dependiendo de factores regionales o de la participación activa de los usuarios (van de Belt et al., 2018; Wittwer et al., 2023).

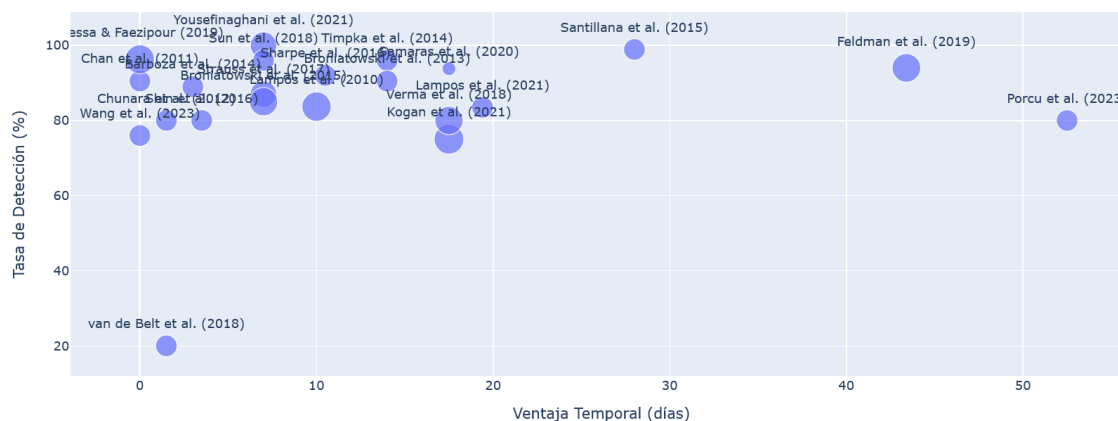
En términos de precisión, en general, la vigilancia digital muestra resultados altos, destacándose correlaciones elevadas (coeficientes mayores de 0.8) y reducción significativa en errores de predicción (RMSE y MAPE bajos), como se observa en estudios como los de Santillana et al. (2015) y Lampos et al. (2015). Sin embargo, la variabilidad geográfica y la penetración tecnológica condicionan estos resultados, reflejando que la precisión y efectividad

no son uniformes y deben adaptarse según la región, tipo de enfermedad y plataforma digital empleada (Kogan et al., 2021).

La figura 7 corresponde a un gráfico de dispersión con burbujas en el que se visualizan distintos estudios sobre vigilancia digital en salud, evaluados con base en tres dimensiones clave. Del análisis del gráfico se destacan varias observaciones. En la esquina superior derecha se ubican los estudios más eficientes, como los de Feldman et al. (2019), Santillana et al. (2015) y Porcu et al. (2023), que combinan una ventaja temporal prolongada (20 a 50 días) con tasas de detección superiores al 85%. Por otro lado, estudios como los de Alessa & Faezipour (2019) o Shin et al. (2016) presentan alta precisión pero con escasa anticipación, lo que limita su utilidad para acciones preventivas. También se evidencian casos con bajo desempeño relativo, como van de Belt et al. (2018), y una alta concentración de estudios con ventajas entre 3 y 15 días, lo que sugiere una barrera técnica en la predicción temprana más allá de dos semanas. Finalmente, modelos con ventaja intermedia pero alta precisión, como los de Kogan et al. (2021) o Samaras et al. (2020), podrían ser valiosos en contextos donde la confiabilidad del pronóstico sea prioritaria.

Figura 7

Relación Entre la Ventaja Temporal y la Tasa de Detección en Estudios de Vigilancia



Aspectos Complementarios y Específicos

El análisis de la Tabla 5 revela aspectos clave en la efectividad de la vigilancia digital en salud pública, centrándose en cuatro dimensiones fundamentales: (1) resolución espacial y granularidad temporal, (2) selección de palabras clave, (3) medición del impacto mediático y (4) consideraciones demográficas y de uso. A continuación, se sintetizan los hallazgos principales sin redundancias.

Tabla 5

Características Complementarias y Variables Contextuales

Autor y año	Resolución espacial y granularidad temporal	Proceso de selección de palabras clave	Medición del impacto mediático	Características demográficas y de uso
Dai et al. (2020)	Regional (Wuhan, China); datos diarios y semanales	Selección manual ('pneumonia', 'SARS')	No se evalúa	No se especifican
Lamos et al. (2010)	Centros urbanos (radio 10 km); agregación diaria y semanal	Selección manual y LASSO	No se mide directamente	5.5 millones de usuarios de Twitter (Reino Unido)
Lamos et al. (2021)	Nacional; datos diarios	19 conjuntos basados en síntomas	Minimiza efecto de pánico en modelo autoregresivo	Aplicación en múltiples países, sin detalles demográficos
van de Belt et al. (2018)	Provincias; datos diarios	Búsquedas Booleanas en Google Trends	No se evalúa explícitamente	Información geográfica de provincias

Autor y año	Resolución espacial y granularidad temporal	Proceso de selección de palabras clave	Medición del impacto mediático	Características demográficas y de uso
Timpka et al. (2014)	Condado; datos diarios	ICD-10 y agrupación en telenursing	Correlación entre cobertura mediática y GFT	Distribución por edad en RIR
McGough et al. (2017)	País; datos semanales	LASSO y regresión penalizada	No se mide, se menciona baja influencia	Perfiles de datos por país, sin demografía
Yousefinaghan i et al. (2021)	Estados/provincias; datos semanales	Listas predefinidas de síntomas	Impacto indirecto al comparar términos preventivos	Geolocalización de tweets por estado/provincia
Wittwer et al. (2023)	Municipios; datos diarios	Cuestionario basado en síntomas COVID-19	Impacto de campañas mediáticas en participación	Tasas de participación y diferencias urbanas
Shin et al. (2016)	Nacional; datos diarios	Términos básicos y extendidos ('MERS', 'hospital')	Reconoce ruido mediático	Datos agregados de búsquedas y tweets
Yan et al. (2017)	Local/global; datos diarios a semanales	Selección por relevancia y especificidad	Evaluación del ruido mediático	Falta de datos detallados sobre usuarios
Strauss et al. (2017)	Nacional; datos semanales	Términos en español sobre dengue	Variación anual de búsqueda vs incidencia	Impacto de penetración de Internet
Chunara et al. (2012)	Departamentos y arrondissements; datos diarios	Búsquedas de 'cholera' y hashtags	Efecto amplificación mediática	Sesgos geográficos y demográficos

Autor y año	Resolución espacial y granularidad temporal	Proceso de selección de palabras clave	Medición del impacto mediático	Características demográficas y de uso
Barboza et al. (2014)	Eventos por país; datos mensuales	Definida por epidemiólogos	Comparación de señales de medios y fuentes oficiales	Distribución de idiomas e impacto regional
Kogan et al. (2021)	Estados; datos diarios	Términos relacionados con COVID-19	Análisis de sesgo en proxies digitales	Diferencias en actividad y adherencia a NPIs
Verma et al. (2018)	Estados; datos semanales	Términos con alta correlación (Google Correlate)	Explosión de búsquedas previas al reporte	Penetración de Internet en Haryana y Chandigarh
Santillana et al. (2015)	Nacional; datos semanales	Términos basados en estudios previos	Captura efectos mediáticos en variación de búsquedas	Nivel agregado nacional sin detalles demográficos
Majumder et al. (2016)	Nacional; datos agregados	Término clave 'Zika' en Google Trends	Comparación de curvas, sin evaluar ruido	Datos agregados sin detalles demográficos
Samaras et al. (2020)	Nacional (Grecia); datos agregados	Términos en griego	Sesgo mediático en búsquedas y tweets	Limitaciones en geolocalización de tweets
Li et al. (2021)	Estatad (EE. UU.); datos diarios	Término clave 'Coronavirus'	Signal ratio como indicador de opinión pública	Filtrado por ubicación en EE. UU.
Feldman et al. (2019)	Global; actualizaciones cada 15 min	Filtrado por GDELT y bases de nombres	Lead time de 43.4 días y 94% cobertura de brotes	Sin características demográficas, solo datos de medios

Autor y año	Resolución espacial y granularidad temporal	Proceso de selección de palabras clave	Medición del impacto mediático	Características demográficas y de uso
Sharpe et al. (2016)	Nacional; datos semanales	Términos implícitos en cada fuente	Evaluación de discrepancias en cambios	Datos agregados sin detalles demográficos
Porcu et al. (2023)	Regiones (Italia); datos semanales	Traducción italiana de síntomas	Volumen de búsqueda como proxy de alerta	Variabilidad de acceso a Internet por región
Sun et al. (2018)	Ciudad; datos semanales	Términos específicos para Boston	Influencia mediática en comparación de métodos	Visitas a urgencias (edad, género, etnia)
Chan et al. (2011)	País; datos semanales/mensuales	Selección por correlación con datos oficiales	No se reporta	No se reporta
Wang et al. (2023)	Nacional y regional; datos semanales	Término 'RSV' o 'RS Virus'	Mitigación de impacto mediático con MEM	No se reportan detalles específicos
Alessa & Faezipour (2019)	Estado (Connecticut); datos semanales	11 palabras clave verificadas	No se mide directamente	No se detallan características demográficas
Broniatowski et al. (2013)	Municipal, regional y nacional; datos semanales	Lista de keywords y modelos previos	Sensibilidad a 'chatter' mediático	Posibles sesgos en subrepresentación de usuarios
Shen et al. (2020)	Nacional y provincial; datos diarios	167 palabras clave por observación diaria	Comparación entre 'sick posts' y otros posts	User pool con composición por edad y género

Autor y año	Resolución espacial y granularidad temporal	Proceso de selección de palabras clave	Medición del impacto mediático	Características demográficas y de uso
Broniatowski et al. (2015)	Municipal (hospital); datos semanales	269 términos de salud filtrados en etapas	Filtrado para reducir chatter mediático	Datos de pacientes pediátricos y adultos

Respecto a la resolución espacial y granularidad temporal, existe una considerable heterogeneidad según el contexto epidemiológico y la disponibilidad de datos digitales. Las escalas varían desde niveles locales o municipales (Wittwer et al., 2023; Broniatowski et al., 2015) hasta nacionales (Shin et al., 2016; Lampos et al., 2021) y globales. Esta variabilidad se explica en gran medida por diferencias en la penetración de internet, lo cual permite resoluciones más precisas en áreas con mayor conectividad digital, como la geolocalización precisa de tuits.

En cuanto al proceso de selección de palabras clave, las estrategias pueden ser manuales o automatizadas. Algunos estudios emplean métodos manuales o listas predefinidas, especialmente para enfermedades específicas como MRSA, cólera o Zika (van de Belt et al., 2018; Majumder et al., 2016), mientras que otros recurren a técnicas automatizadas o herramientas como Google Correlate (Verma et al., 2018). Esta selección es determinante, ya que afecta directamente la calidad y representatividad de las señales digitales.

El impacto mediático también juega un papel relevante. Algunos estudios consideran explícitamente la influencia mediática, utilizando modelos para mitigar efectos del pánico o la sobreexposición informativa (Lampos et al., 2021; Kogan et al., 2021). Otras investigaciones integran datos de cobertura mediática local o internacional para ajustar las señales digitales ante potenciales distorsiones informativas.

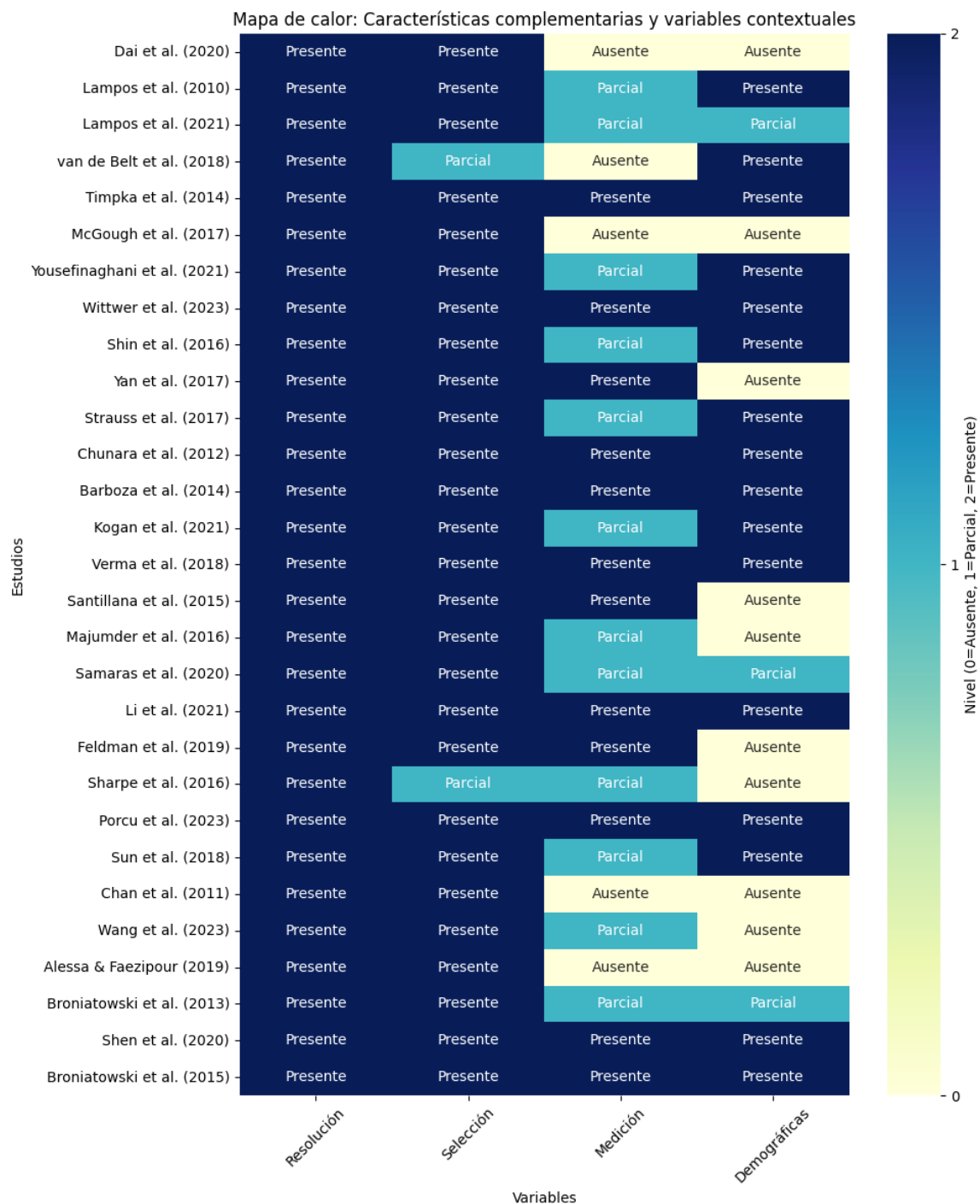
La dimensión demográfica y de uso revela limitaciones en cuanto a la representación detallada por edad, género o nivel socioeconómico, observándose pocas excepciones como Timpka et al. (2014), que utiliza índices específicos para grupos etarios. Además, varios estudios mencionan explícitamente posibles sesgos derivados de la geolocalización y el acceso diferenciado a tecnologías digitales, factores que pueden afectar la representatividad y precisión de las señales obtenidas (Dai et al., 2020; Lampos et al., 2021).

La figura 8 evidencia que, mientras la mayoría de los estudios abordan adecuadamente la resolución espacio-temporal y la selección de palabras clave, aspectos como el impacto mediático y la demografía suelen tratarse de forma parcial o ser ignorados. Esto refleja un sesgo frecuente al subestimar factores humanos y comunicacionales, a pesar de su relevancia para comprender la propagación y percepción de las enfermedades. La Figura 8 evidencia una cobertura desigual de las variables contextuales y complementarias en la vigilancia digital. Mientras que la resolución y la selección de palabras clave son ampliamente consideradas, la medición del impacto mediático y la inclusión de variables demográficas aún representan puntos débiles. Esta asimetría metodológica puede inducir sesgos importantes, especialmente al interpretar la propagación de enfermedades o al diseñar intervenciones basadas en datos digitales.

Por tanto, los hallazgos sugieren que para optimizar el valor predictivo y la aplicabilidad de los sistemas de vigilancia digital, es esencial integrar de forma más sistemática y rigurosa todas estas dimensiones. Su consideración conjunta permite una aproximación más robusta, sensible y representativa a los fenómenos de salud pública en contextos digitales.

Figura 8

Mapa de Calor de las Características Complementarias y Variables Contextuales



Discusión

Principales Hallazgos y Evolución Metodológica

La presente revisión evidencia una notable diversidad metodológica en los estudios de vigilancia digital en salud pública, abarcando desde análisis empíricos comparativos y estudios observacionales hasta investigaciones avanzadas basadas en modelado matemático y técnicas de *machine learning*. Se observa una evolución clara hacia enfoques cada vez más sofisticados que incorporan innovaciones tales como el aprendizaje por transferencia, análisis bayesiano y modelado de series temporales. Estas técnicas modernas no solo permiten mejorar la precisión en la identificación de brotes epidemiológicos, sino que también facilitan la comprensión de relaciones temporales y causales entre las señales digitales y los datos oficiales.

Los estudios revisados se extienden en el tiempo, partiendo de investigaciones pioneras sobre el dengue en 2003 hasta análisis recientes centrados en emergentes amenazas como la COVID-19. Esta variabilidad temporal responde tanto a brotes prolongados (como la influenza) como a eventos de rápida evolución (por ejemplo, Zika y COVID-19), lo cual evidencia la capacidad de adaptación de las metodologías analíticas según las características del patógeno. Además, desde una perspectiva geográfica se destaca la aplicación de estos métodos en contextos locales, nacionales e internacionales, lo que permite evaluar la influencia de factores contextuales, tales como la conectividad digital y la densidad poblacional, sobre la eficacia de la vigilancia digital.

Integración de Fuentes Digitales con Sistemas Tradicionales

Una característica recurrente en los estudios es el uso combinado de datos provenientes de motores de búsqueda (como *Google Trends*) y redes sociales (principalmente “X”, anteriormente Twitter). Estos datos se validan sistemáticamente con reportes oficiales de

instituciones reconocidas, tales como el CDC, la OMS y ministerios de salud. Esta integración bidireccional no solo otorga mayor credibilidad a los métodos digitales, sino que también refuerza su utilidad práctica al proporcionar alertas tempranas que, en ciertos casos, pueden anticipar los brotes desde días hasta incluso semanas antes que los sistemas convencionales.

Capacidades de Detección Temprana y Precisión Predictiva

Uno de los hallazgos más relevantes es la capacidad de los enfoques digitales para detectar anomalías y tendencias epidemiológicas de manera anticipada. Los estudios indican que estas metodologías pueden ofrecer una ventaja temporal significativa, con alertas que en algunos casos sobrepasan en hasta 43 días la notificación oficial. Los altos coeficientes de correlación y las reducciones en errores predictivos (medidos a través de métricas como RMSE y MAE) avalan la capacidad de estos modelos para anticipar brotes de forma precisa. No obstante, se ha observado que tanto la precisión como la ventaja temporal dependen del tipo de enfermedad, la fuente de datos empleada y la metodología aplicada, lo que subraya la importancia de adaptar los modelos analíticos a las particularidades contextuales y epidemiológicas de cada región.

Consideraciones Complementarias: Resolución, Selección de Palabras Clave, Impacto Mediático y Variables Demográficas

El éxito de la vigilancia digital no depende únicamente de la robustez de los modelos estadísticos y de aprendizaje automático, sino también de la calidad y el carácter contextual de los datos utilizados. Aspectos como la resolución espacial y la granularidad temporal permiten detectar brotes con mayor precisión, abarcando escalas que van desde el nivel municipal hasta el nacional. De igual forma, la adecuada selección de palabras clave, ya sea mediante métodos manuales o automatizados, resulta determinante para extraer señales relevantes y minimizar el ruido digital.

Por otro lado, el impacto mediático puede influir en la dinámica de los datos, distorsionando temporalmente las señales debido a la sobreexposición informativa o el efecto pánico. Por ello, diversos estudios han implementado estrategias para mitigar estos efectos, ajustando sus modelos en función de la cobertura mediática. Finalmente, la consideración de variables demográficas (como edad, género y acceso a tecnologías) es crucial para enfrentar posibles sesgos, garantizando que las predicciones sean representativas y precisas, tal como enfatizan investigaciones recientes (Shakeri Hossein Abad et al., 2021).

Comparación con la Literatura Previa

Los hallazgos de esta revisión concuerdan en parte con investigaciones anteriores en el campo de la vigilancia digital, que han resaltado el predominio de plataformas como Google y Twitter para la detección temprana de brotes (ver tabla 6). El estudio de Al-Kenane et al. (2024) (Al-Kenane et al., 2024) se enfoca específicamente en el contexto de Kuwait, utilizando Google Trends en inglés y árabe para explorar la relación entre la actividad de búsqueda y la respuesta gubernamental ante brotes. A diferencia del presente trabajo, Al-Kenane et al. (Al-Kenane et al., 2024) destacan particularmente la influencia de variables culturales y psicológicas mediante un enfoque bilingüe, aportando una dimensión importante pero limitada geográficamente. Aunque su metodología es innovadora, la generalización de sus hallazgos es más restringida debido a las características específicas del contexto estudiado.

Tabla 6*Comparación de Resultados con Otros Estudios*

Autores/Estudio	Objetivo / Alcance	Metodología / Técnicas	Fuentes de Datos	Resultados Clave	Impacto / Contexto	Ventajas / Desventajas
Este estudio (2025)	Evaluar el uso de redes sociales y fuentes digitales para la detección temprana de brotes de enfermedades infecciosas.	Análisis retrospectivo y predictivo; uso de aprendizaje automático, correlaciones y análisis de series temporales.	Twitter, Google Trends, foros de salud, bases de datos de noticias, registros epidemiológicos.	Anticipación de brotes hasta varias semanas antes; correlaciones altas con reportes oficiales.	Impacto del contexto mediático y penetración digital en la calidad de los datos.	<ul style="list-style-type: none"> ✓ Integración de múltiples fuentes digitales; validación con datos oficiales. ✗ Variabilidad en la representatividad de los datos según la región y el acceso digital.
(Al-Kenane et al., 2024)	Relación entre búsquedas en Google Trends y respuesta gubernamental en Kuwait.	Análisis de series temporales; Pearson y bootstrap.	Google Trends (inglés y árabe).	Alta correlación ($R \sim 0.71$); anticipa cambios en políticas.	Incorpora análisis bilingüe y efectos de medidas gubernamentales.	<ul style="list-style-type: none"> ✓ Enfoque innovador y robusto. ✗ Limitado a Kuwait y variables psicológicas.

Autores/Estudio	Objetivo / Alcance	Metodología / Técnicas	Fuentes de Datos	Resultados Clave	Impacto / Contexto	Ventajas / Desventajas
(Melo et al., 2024)	Evaluar herramientas digitales para la vigilancia de arbovirus y detección temprana.	Revisión con análisis comparativo; ANOVA y correlaciones.	Google Trends, Twitter, apps, redes sociales y datos oficiales.	Detección anticipada (días a semanas); alta precisión en predicción de brotes.	Considera influencia mediática y resolución espacial de datos.	✓ Comparación integral de herramientas y contextos. ✗ Alta variabilidad entre estudios.
(Jia et al., 2023)	Revisión de innovaciones tecnológicas (AI, GIS, digital twin) en la vigilancia epidemiológica.	Síntesis en Annual Review of Public Health.	Datos geoespaciales, EHRs, big data, notificación electrónica.	Mejora en precisión, oportunidad y detección en tiempo real.	Impacto de smart devices y evolución digital en salud pública.	✓ Destaca avances clave en vigilancia. ✗ Heterogeneidad en estudios; requiere integración con otros sistemas.
(Zhao et al., 2021)	Análisis ético de vigilancia digital en enfermedades infecciosas.	Revisión sistemática con enfoque teórico en privacidad y derechos civiles.	Datos de big data, EHRs y vigilancia digital.	Evalúa riesgos éticos vs. beneficios; correlación con reportes oficiales.	Resalta necesidad de equilibrio entre vigilancia y	✓ Marco teórico sólido sobre ética en vigilancia digital. ✗ No aborda métricas operativas.

Autores/Estudio	Objetivo / Alcance	Metodología / Técnicas	Fuentes de Datos	Resultados Clave	Impacto / Contexto	Ventajas / Desventajas
(Salathé et al., 2012)	Impacto del big data y redes sociales en epidemiología digital.	Revisión narrativa y "Editors' Outlook".	Redes sociales, telefonía móvil, búsquedas en línea.	Reducción del tiempo de detección de brotes.	Privacidad. Potencial de alertas tempranas vs. desafíos técnicos y sesgos.	<ul style="list-style-type: none"> ✓ Pionero en digital epidemiology. ✗ Falta de métricas de error detalladas.

Por otra parte, Melo et al. (2024) (Melo et al., 2024) comparten objetivos similares al presente estudio, ya que también evalúan múltiples herramientas digitales como Google Trends, Twitter y aplicaciones móviles en la vigilancia de arbovirus. Su revisión comparativa, basada en métodos estadísticos como ANOVA y correlaciones, corrobora las ventajas identificadas en cuanto a la detección anticipada y precisión predictiva. Sin embargo, Melo et al. también subrayan una alta variabilidad metodológica entre estudios, un desafío compartido con nuestro estudio actual, enfatizando la necesidad de una mayor estandarización en el campo.

El trabajo de Peng Jia et al. (2023) (Jia et al., 2023) ofrece una perspectiva complementaria al enfocarse en innovaciones tecnológicas avanzadas como la inteligencia artificial, GIS y digital twins, demostrando mejoras significativas en términos de precisión y detección en tiempo real. Si bien no se centra exclusivamente en arbovirus, este estudio amplía el contexto de la vigilancia digital, destacando cómo los dispositivos inteligentes y la evolución

tecnológica potencian la respuesta epidemiológica. Aun así, reconoce la heterogeneidad metodológica como una barrera persistente, similar a lo encontrado en nuestro análisis.

Zhao et al. (2021) (Zhao et al., 2021), por su parte, aportan una dimensión crítica y complementaria al analizar aspectos éticos relacionados con la vigilancia digital en salud pública. Su enfoque se distancia del operativo para abordar temas cruciales como la privacidad y la protección de derechos civiles, factores esenciales para asegurar la aceptación social y sostenibilidad a largo plazo de estos sistemas. Aunque su análisis carece de métricas operativas concretas, destaca la importancia del equilibrio entre eficiencia técnica y responsabilidad ética, algo que nuestro estudio reconoce indirectamente al señalar el impacto del contexto mediático y digital.

El trabajo pionero de Salathé et al. (2012) (Salathé et al., 2012) sienta las bases conceptuales de la epidemiología digital, resaltando el potencial inicial de las redes sociales y Big Data para reducir los tiempos de respuesta en la detección de brotes. Aunque carece de métricas detalladas como las proporcionadas por estudios más recientes, su contribución inicial facilita entender cómo el campo ha evolucionado hacia enfoques metodológicos más rigurosos y cuantitativos, como se ejemplifica en este estudio actual.

Al comparar el estudio actual con el de Shakeri et al. (2021) (Shakeri Hossein Abad et al., 2021), se encuentra una amplia diversidad metodológica y temporal en los enfoques analizados, desde investigaciones pioneras sobre enfermedades como el dengue hasta estudios recientes enfocados en COVID-19. Se destaca especialmente el creciente uso de técnicas avanzadas como el aprendizaje por transferencia, análisis bayesiano y modelos de series temporales, principalmente en enfermedades respiratorias y virales. Asimismo, se resalta la importancia de adaptar estas metodologías a contextos regionales específicos y de integrar

eficazmente fuentes digitales como Google Trends y redes sociales con los sistemas tradicionales de vigilancia epidemiológica.

El *scoping review* de Shakeri et al. (Shakeri Hossein Abad et al., 2021) ofrece una perspectiva más extensa y detallada sobre el uso de plataformas digitales en la vigilancia en salud pública, cubriendo no solo enfermedades infecciosas, sino también áreas como la salud mental y enfermedades crónicas. Este estudio destaca también limitaciones significativas, como sesgos metodológicos derivados de la selección de palabras clave y la escasa aplicación práctica de los resultados en acciones concretas de salud pública. Ambos trabajos coinciden en subrayar la necesidad de mejorar continuamente las metodologías y fortalecer la integración entre los datos digitales y los sistemas tradicionales, para maximizar el impacto de las intervenciones en salud pública. Con relación a lo hallado, es importante enfatizar que la complementariedad entre estos enfoques analíticos y la vigilancia tradicional permite mejorar la capacidad de respuesta ante emergencias sanitarias. Esto implica no solo avanzar tecnológicamente, sino también considerar factores socioculturales y operativos que faciliten una mayor efectividad en las intervenciones sanitarias derivadas de la vigilancia digital.

En síntesis, al comparar explícitamente nuestro enfoque con el de otros estudios, se destaca que:

Novedad Metodológica: Nuestro estudio integra múltiples fuentes digitales y aplica técnicas avanzadas de *machine learning* y modelado de series temporales, superando así el alcance limitado de investigaciones centradas en contextos geográficos o tecnológicas específicas.

Aplicación Práctica: La validación sistemática contra datos oficiales y la consideración del impacto mediático y las variables demográficas refuerzan la utilidad operativa del enfoque, permitiendo alertas tempranas de brotes con ventajas de hasta varias semanas.

Contextualización de Limitaciones: Mientras estudios previos destacan limitaciones en la generalización y la heterogeneidad metodológica, nuestro trabajo vincula estas limitaciones con evidencias empíricas previas y propone estrategias específicas (como la mejora en la selección de palabras clave y la integración de múltiples fuentes) para superarlas en futuras investigaciones.

Esta comparación explícita subraya cómo nuestro enfoque contribuye a la evolución del campo de la vigilancia digital en salud pública, ofreciendo una metodología más completa y adaptable a diversos contextos epidemiológicos y operativos.

Consideraciones Metodológicas

El análisis de los métodos y técnicas empleadas en la vigilancia digital revela un flujo metodológico compartido, que abarca desde la recolección de datos hasta su evaluación, y destaca cómo la incorporación de técnicas avanzadas mejora sustancialmente la detección temprana de brotes. En la fase inicial, se recolectan datos de diversas fuentes digitales, motores de búsqueda, redes sociales, plataformas especializadas y datos de movilidad, seguidos de un riguroso preprocesamiento que incluye normalización y suavizado, garantizando la claridad y calidad de las señales epidemiológicas.

Discusión de Técnicas Específicas

Uno de los aspectos metodológicos más innovadores es el uso del aprendizaje por transferencia y el análisis bayesiano. Por ejemplo, algunos estudios (como el de Lampos et al., 2021(Lampos et al., 2021), descrito en la Tabla 3) han implementado modelos *unsupervised*

combinados con técnicas de *transfer learning*. Estos métodos permiten aprovechar información previamente aprendida de grandes volúmenes de datos, facilitando la detección de patrones sutiles en contextos nuevos y mejorando así la capacidad predictiva en escenarios con datos limitados. De igual manera, el empleo de algoritmos bayesianos, utilizados por autores como Sharpe et al. (2016) (Sharpe et al., 2016) y Yan et al. (2017) (Yan et al., 2017), ha demostrado ser eficaz para identificar puntos de cambio en series temporales. Estos algoritmos permiten modelar la incertidumbre de forma más precisa y han mostrado, en ocasiones, mejoras en la correlación con datos oficiales, lo que se traduce en alertas más confiables para la detección temprana.

Claridad en el Proceso de Integración

La integración de datos digitales con fuentes oficiales es otro pilar fundamental que potencia la precisión predictiva de la vigilancia digital. Por ejemplo, el estudio de Timpka et al. (2014) (Timpka et al., 2014) combinó señales de Google Flu Trends con datos clínicos y de laboratorio, alcanzando coeficientes de correlación cercanos a 0.96, lo que evidencia la sinergia entre ambos tipos de fuentes. Esta integración no solo refuerza la validez de los modelos predictivos, sino que permite mitigar los sesgos inherentes a la recolección exclusiva de datos digitales. Un ejemplo adicional lo encontramos en el trabajo de Yousefinaghani et al. (2021) (Yousefinaghani et al., 2021), donde el uso simultáneo de la API de Twitter y Google Trends, contrastado con reportes oficiales (como los del Johns Hopkins COVID-19), resultó en una mejora notable en la precisión de las predicciones, con métricas de error reducidas (RMSE y MAE) y altos coeficientes de correlación (por encima del 75%).

Selección de Palabras Clave y Gestión del Impacto Mediático

Paralelamente, la adecuada selección de palabras clave, aplicada a través de métodos que van desde la identificación manual hasta herramientas automatizadas como Google Correlate es determinante para extraer señales relevantes y minimizar el ruido digital. Estrategias específicas, descritas en estudios como los de Shakeri Hossein Abad et al. (2021) (Shakeri Hossein Abad et al., 2021), han permitido reducir la interferencia de la sobreexposición mediática, ajustando las predicciones en función de variaciones en la atención informativa y, a la vez, garantizando la representatividad de las señales. Este enfoque contribuye a una mejor interpretación de los datos, optimizando la detección temprana y la respuesta operativa en salud pública.

En conjunto, la combinación de técnicas avanzadas, como el aprendizaje por transferencia y análisis bayesiano, con un proceso de integración riguroso entre datos digitales y fuentes oficiales, se traduce en mejoras significativas en la predicción de brotes. Los ejemplos concretos extraídos de la Tabla 3 respaldan que estas estrategias no solo incrementan la capacidad de alerta temprana (con ventajas de varios días a semanas), sino que también reducen los errores predictivos, posicionando a estos métodos como herramientas complementarias esenciales para la vigilancia epidemiológica.

Limitaciones de los Estudios

A pesar de los avances significativos en vigilancia digital presentados por los estudios analizados, se identifican importantes limitaciones metodológicas que merecen una discusión más profunda. En primer lugar, existe una notable heterogeneidad en la resolución espacial y granularidad temporal, que abarca desde escalas muy locales, como el nivel municipal, hasta análisis globales. Esta variabilidad complica considerablemente la comparabilidad y generalización de resultados, ya que las señales epidemiológicas pueden presentar patrones

distintos dependiendo del contexto geográfico y temporal evaluado. Por tanto, sería necesario establecer estándares mínimos que permitan un análisis comparativo más robusto, facilitando la interpretación de resultados entre estudios diversos (Dhewantara et al., 2019; Shakeri Hossein Abad et al., 2021; Shaweno et al., 2018).

Asimismo, otro punto crítico es la selección de palabras clave, procedimiento fundamental para garantizar la calidad y representatividad de las señales digitales. Actualmente, muchos estudios recurren a metodologías manuales o semiautomáticas para determinar estas palabras clave, lo cual podría no capturar plenamente la complejidad y evolución del lenguaje digital, especialmente en contextos culturales o lingüísticos heterogéneos. Esto podría ocasionar sesgos en la identificación temprana de brotes, subestimando o sobreestimando ciertos términos en función de criterios subjetivos o limitaciones técnicas. Por ende, avanzar hacia métodos más sofisticados, como modelos lingüísticos basados en aprendizaje automático, podría permitir una selección más dinámica, adaptativa y precisa de términos clave (Feldman et al., 2019; Shakeri Hossein Abad et al., 2021; Sulaiman et al., 2019).

Otro desafío sustancial reside en la insuficiente consideración de variables demográficas y tecnológicas en la mayoría de los estudios. La falta de análisis detallados sobre la composición demográfica de los usuarios digitales, incluyendo factores como edad, género, educación, nivel socioeconómico y ubicación, puede limitar la representatividad de las señales epidemiológicas obtenidas. Además, la brecha digital, marcada por disparidades en el acceso y uso de tecnologías entre áreas urbanas y rurales o entre países de diferente nivel socioeconómico, puede sesgar significativamente los resultados, dificultando su generalización y aplicabilidad universal. Por tanto, integrar análisis demográficos detallados y estudios sobre la penetración tecnológica

permitiría interpretar mejor los datos y ajustar las conclusiones a realidades específicas (Cho et al., 2022; Ragnedda et al., 2024).

Por último, aunque la mayoría de las investigaciones valida sus hallazgos mediante comparación con datos oficiales, la integración sistemática de estas metodologías digitales con sistemas tradicionales de vigilancia epidemiológica sigue siendo limitada. Aún se requiere desarrollar protocolos claros y estandarizados que faciliten la combinación efectiva y transparente de ambas fuentes de datos. En este sentido, estudios multicéntricos coordinados internacionalmente podrían proporcionar la base necesaria para crear estándares consensuados que aseguren la validez externa y permitan aplicar los resultados de manera confiable y generalizada a nivel global (Aiello et al., 2020; Shakeri Hossein Abad et al., 2021).

Implicaciones Clínicas y Recomendaciones para Investigaciones Futuras

La vigilancia digital se presenta como un complemento prometedor para los sistemas tradicionales de vigilancia epidemiológica, al posibilitar la detección temprana de brotes y el análisis casi en tiempo real de las tendencias epidemiológicas. Sin embargo, para aprovechar plenamente este potencial y superar los desafíos actuales, es necesario implementar protocolos estandarizados y estrategias de integración más robustas, combinados con un enfoque interdisciplinario que reúna a técnicos, epidemiólogos y científicos sociales (Fallatah & Adekola, 2024; Kostkova et al., 2021).

Recomendaciones Claras y Accionables

Estandarización de Protocolos: Es imperativo desarrollar estudios multicéntricos que utilicen métodos uniformes en todas las fases del análisis. Esto incluiría la adopción de protocolos estandarizados para la selección de palabras clave, ya sea mediante métodos manuales o automatizados, y para el preprocesamiento de datos (normalización, suavizado, filtrado). Por

ejemplo, se podrían establecer directrices comunes que integren métricas de validación similares a las empleadas en los trabajos de Brancato et al. (2024) o Jacobson et al. (2024), facilitando así la comparación de resultados entre diferentes estudios y contextos geográficos (Brancato et al., 2024; Jacobson et al., 2024).

Integración Sistemática del Análisis Demográfico: Se recomienda incorporar de forma sistemática variables geográficas, socioeconómicas y demográficas en los modelos predictivos. A corto plazo, podrían desarrollarse plataformas digitales que integren estas variables mediante la vinculación de datos abiertos de censos y registros oficiales con datos recolectados de redes sociales y motores de búsqueda. Este enfoque permitiría mejorar la representatividad y precisión de las señales epidemiológicas, reduciendo sesgos derivados de las variaciones regionales en el acceso y uso de tecnologías digitales (Ragnedda et al., 2024; Shakeri Hossein Abad et al., 2021).

Uso de Tecnologías Emergentes: La adopción de herramientas de inteligencia artificial y técnicas avanzadas como el análisis metagenómico podría optimizar la detección de brotes. Es recomendable implementar técnicas de clasificación y filtrado automático para analizar grandes volúmenes de datos, lo que incrementaría la sensibilidad y capacidad predictiva de los sistemas digitales, especialmente cuando se combinan con fuentes tradicionales de vigilancia (Syrowatka, 2021).

Evaluaciones Longitudinales: Se sugiere la realización de estudios de seguimiento a largo plazo que evalúen la eficacia, estabilidad y rentabilidad operativa de los sistemas digitales en diversos contextos epidemiológicos. Este enfoque no solo proporcionaría evidencia robusta sobre la sostenibilidad de la vigilancia digital, sino que también ayudaría a ajustar y perfeccionar los modelos predictivos a lo largo del tiempo.

Perspectiva Interdisciplinaria

Para abordar la complejidad del análisis digital en salud pública, es fundamental fomentar la colaboración entre técnicos, epidemiólogos y científicos sociales. Esta colaboración interdisciplinaria permitiría:

Integrar conocimientos técnicos y contextuales: Los expertos en tecnología pueden optimizar algoritmos y modelos predictivos, mientras que los epidemiólogos aportan su conocimiento de la dinámica de las enfermedades y los científicos sociales ofrecen perspectivas sobre factores culturales, demográficos y de comportamiento, que son esenciales para interpretar las señales digitales.

Diseñar intervenciones adaptadas y equitativas: Este enfoque colaborativo facilitará el diseño de intervenciones de salud pública más precisas y adaptadas a las realidades locales, maximizando el impacto en la prevención y control de brotes.

Desarrollar soluciones holísticas: Al combinar habilidades y conocimientos de diversas disciplinas, se pueden desarrollar soluciones integrales que aborden tanto aspectos operativos como éticos, garantizando que la vigilancia digital se implemente de forma responsable y con altos estándares de eficacia.

En conjunto, aunque la evidencia actual respalda ampliamente la capacidad predictiva de las herramientas digitales en vigilancia epidemiológica, persisten desafíos metodológicos y operativos que dificultan su generalización clínica. Por tanto, se recomienda adoptar un enfoque interdisciplinario y colaborativo que combine innovaciones tecnológicas avanzadas con diseños metodológicos rigurosos, fortaleciendo así la capacidad de respuesta de los sistemas de salud pública frente a emergencias epidemiológicas.

Conclusiones

La evidencia recopilada en esta revisión demuestra que la integración de diversas fuentes digitales, como motores de búsqueda, redes sociales y bases de datos especializadas, permite anticipar brotes de enfermedades infecciosas con una ventaja temporal considerable respecto a los sistemas tradicionales de vigilancia. Estos métodos, que utilizan técnicas avanzadas de análisis estadístico y aprendizaje automático, logran altos niveles de correlación y precisión al cotejar sus hallazgos con datos oficiales. Sin embargo, la heterogeneidad en la resolución espacial y temporal de los datos, la variabilidad en la selección y procesamiento de términos clave, la gestión del ruido mediático y la limitada incorporación de variables demográficas y de uso constituyen desafíos que afectan la generalización clínica y la validez externa de los resultados.

Por ello, para optimizar la implementación práctica de la vigilancia digital en salud pública es esencial la estandarización de protocolos metodológicos que unifiquen la selección de palabras clave y el preprocesamiento de datos. Además, se recomienda profundizar en el análisis demográfico y en la integración de tecnologías emergentes, como la inteligencia artificial, para reducir sesgos y mejorar la precisión y representatividad de los sistemas de detección temprana. Esta combinación de estrategias no solo facilitará la toma de decisiones informadas en tiempo real, sino que también sentará las bases para el desarrollo de enfoques más robustos y eficientes en la lucha contra las emergencias epidemiológicas.

Bibliografia

- Aiello, A. E., Renson, A., & Zivich, P. N. (2020). Social Media- and Internet-Based Disease Surveillance for Public Health. *Annual Review of Public Health, 41*, 101-118.
<https://doi.org/10.1146/annurev-publhealth-040119-094402>
- Al-Kenane, K., Boy, F., Alsaber, A., Nafea, R., & AlMutairi, S. (2024). Digital epidemiology of high-frequency search listening trends for the surveillance of subjective well-being during COVID-19 pandemic. *Frontiers in Psychology, 15*, 1442303.
<https://doi.org/10.3389/fpsyg.2024.1442303>
- Araújo, D. H. M. de, Carvalho, E. A. de, Carvalho, P. V. R. de, & Gomes, J. O. (2020). Social networks applied to Dengue, H1N1, and zika epidemics: An integrative literature review. *Work*. <https://api.semanticscholar.org/CorpusID:226288169>
- Barros, J. M., Duggan, J., & Rebholz-Schuhmann, D. (2020). The Application of Internet-Based Sources for Public Health Surveillance (Infoveillance): Systematic Review. *Journal of Medical Internet Research, 22*(3), e13680. <https://doi.org/10.2196/13680>
- Brancato, V., Esposito, G., Coppola, L., Cavaliere, C., Mirabelli, P., Scapicchio, C., Borgheresi, R., Neri, E., Salvatore, M., & Aiello, M. (2024). Standardizing digital biobanks: Integrating imaging, genomic, and clinical data for precision medicine. *Journal of Translational Medicine, 22*(1), 136. <https://doi.org/10.1186/s12967-024-04891-8>
- Bruneau, G. A. (2020). *The ethical imperatives of the COVID-19 pandemic: A review from data ethics*.
- Budd, J., Miller, B. S., Manning, E. M., Lampos, V., Zhuang, M., Edelstein, M., Rees, G., Emery, V. C., Stevens, M. M., Keegan, N., Short, M. J., Pillay, D., Manley, E., Cox, I. J., Heymann, D., Johnson, A. M., & McKendry, R. A. (2020). Digital technologies in the

- public-health response to COVID-19. *Nature Medicine*, 26(8), 1183-1192.
<https://doi.org/10.1038/s41591-020-1011-4>
- Chen, T., Rosen, R., Grace, W., & Alden, D. (2022). Case Report: A Case of Adult Nesidioblastosis. *HPB*, 24, S328. <https://doi.org/10.1016/j.hpb.2022.05.694>
- Cho, P. J., Yi, J., Ho, E., Shandhi, M. M. H., Dinh, Y., Patil, A., Martin, L., Singh, G., Bent, B., Ginsburg, G., Smuck, M., Woods, C., Shaw, R., & Dunn, J. (2022). Demographic Imbalances Resulting From the Bring-Your-Own-Device Study Design. *JMIR mHealth and uHealth*, 10(4), e29510. <https://doi.org/10.2196/29510>
- Dai, Y., & Wang, J. (2020). Identifying the outbreak signal of COVID-19 before the response of the traditional disease monitoring system. *PLOS Neglected Tropical Diseases*, 14(10), e0008758. <https://doi.org/10.1371/journal.pntd.0008758>
- De La Paz Lascano, M. (2022). Marco normativo para aplicaciones digitales en salud en la Argentina: Una necesidad a resolver. *Revista Chilena de Derecho*, 48(3), 209-216.
<https://doi.org/10.7764/r.483.10>
- Dhewantara, P. W., Lau, C. L., Allan, K. J., Hu, W., Zhang, W., Mamun, A. A., & Soares Magalhães, R. J. (2019). Spatial epidemiological approaches to inform leptospirosis surveillance and control: A systematic review and critical appraisal of methods. *Zoonoses and Public Health*, 66(2), 185-206. <https://doi.org/10.1111/zph.12549>
- Dueñas, D., Daza, J., & Liscano, Y. (2023). Coinfections and Superinfections Associated with COVID-19 in Colombia: A Narrative Review. *Medicina*, 59(7), 1336.
<https://doi.org/10.3390/medicina59071336>

- Escarcena, I. L. (2021). Desafíos éticos en el estudio de comunidades digitales: Una propuesta basada en un caso empírico. *Cuadernos.info*.
<https://api.semanticscholar.org/CorpusID:236748027>
- Fallatah, D., & Adekola, H. A. (2024). Digital epidemiology: Harnessing big data for early detection and monitoring of viral outbreaks. *Infection Prevention in Practice*, 6.
<https://api.semanticscholar.org/CorpusID:270864612>
- Feldman, J., Thomas-Bachli, A., Forsyth, J., Patel, Z. H., & Khan, K. (2019). Development of a global infectious disease activity database using natural language processing, machine learning, and human expertise. *Journal of the American Medical Informatics Association*, 26(11), 1355-1359. <https://doi.org/10.1093/jamia/ocz112>
- Groseclose, S. L., & Buckeridge, D. L. (2017). Public Health Surveillance Systems: Recent Advances in Their Use and Evaluation. *Annual Review of Public Health*, 38(1), 57-79.
<https://doi.org/10.1146/annurev-publhealth-031816-044348>
- Haddaway, N. R., Page, M. J., Pritchard, C. C., & McGuinness, L. A. (2022). PRISMA2020: An R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and Open Synthesis. *Campbell Systematic Reviews*, 18(2), e1230. <https://doi.org/10.1002/cl2.1230>
- Hunter, R. F., Gough, A., O’Kane, N., McKeown, G., Fitzpatrick, A., Walker, T., McKinley, M. C., Lee, M., & Kee, F. (2018). Ethical Issues in Social Media Research for Public Health. *American journal of public health*, 108 3, 343-348.
- Hussain-Alkhateeb, L., Kroeger, A., Olliaro, P., Rocklöv, J., Sewe, M. O., Tejada, G., Benitez, D., Gill, B., Hakim, S. L., Gomes Carvalho, R., Bowman, L., & Petzold, M. (2018). Early warning and response system (EWARS) for dengue outbreaks: Recent advancements

towards widespread applications in critical settings. *PLOS ONE*, 13(5), e0196811.

<https://doi.org/10.1371/journal.pone.0196811>

Ibrahim, N. K. (2020). Epidemiologic surveillance for controlling Covid-19 pandemic: Types, challenges and implications. *Journal of Infection and Public Health*, 13, 1630-1638.

Jacobson, L. P., Parker, C. B., Cella, D., Mroczek, D. K., Lester, B. M., on behalf of program collaborators for Environmental influences on Child Health Outcomes, Smith, P. B., Newby, K. L., Catellier, D. J., Gershon, R., & Cella, D. (2024). Approaches to protocol standardization and data harmonization in the ECHO-wide cohort study. *Pediatric Research*, 95(7), 1726-1733. <https://doi.org/10.1038/s41390-024-03039-0>

Jia, P., Liu, S., & Yang, S. (2023). Innovations in Public Health Surveillance for Emerging Infections. *Annual Review of Public Health*, 44(1), 55-74.

<https://doi.org/10.1146/annurev-publhealth-051920-093141>

Kogan, N. E., Clemente, L., Liataud, P., Kaashoek, J., Link, N. B., Nguyen, A. T., Lu, F. S., Huybers, P., Resch, B., Havas, C., Petutschnig, A., Davis, J., Chinazzi, M., Mustafa, B., Hanage, W. P., Vespignani, A., & Santillana, M. (2021). An early warning approach to monitor COVID-19 activity with multiple digital traces in near real time. *Science Advances*, 7(10), eabd6989. <https://doi.org/10.1126/sciadv.abd6989>

Kostkova, P., Saigí-Rubió, F., Eguia, H., Borbolla, D., Verschuuren, M., Hamilton, C., Azzopardi-Muscat, N., & Novillo-Ortiz, D. (2021). Data and Digital Solutions to Support Surveillance Strategies in the Context of the COVID-19 Pandemic. *Frontiers in Digital Health*, 3, 707902. <https://doi.org/10.3389/fdgth.2021.707902>

- Lamos, V., Majumder, M. S., Yom-Tov, E., Edelstein, M., Moura, S., Hamada, Y., Rangaka, M. X., McKendry, R. A., & Cox, I. J. (2021). Tracking COVID-19 using online search. *Npj Digital Medicine*, 4(1), 17. <https://doi.org/10.1038/s41746-021-00384-w>
- Luo, S., Yin, L., Liu, X., & Wang, X. (2025). Advances in Virus Biorecognition and Detection Techniques for the Surveillance and Prevention of Infectious Diseases. *Biosensors*, 15(3), 198. <https://doi.org/10.3390/bios15030198>
- MacIntyre, C. R., Lim, S., Gurdasani, D., Miranda, M., Metcalf, D., Quigley, A. L., Hutchinson, D., Burr, A., & Heslop, D. J. (2023). Early detection of emerging infectious diseases—Implications for vaccine development. *Vaccine*. <https://api.semanticscholar.org/CorpusID:259059871>
- Mavragani, A. (2020). Infodemiology and Infoveillance: Scoping Review. *Journal of Medical Internet Research*, 22(4), e16206. <https://doi.org/10.2196/16206>
- McClymont, H., Lambert, S. B., Barr, I., Vardoulakis, S., Bambrick, H., & Hu, W. (2024). Internet-based Surveillance Systems and Infectious Diseases Prediction: An Updated Review of the Last 10 Years and Lessons from the COVID-19 Pandemic. *Journal of Epidemiology and Global Health*, 14(3), 645-657. <https://doi.org/10.1007/s44197-024-00272-y>
- Meckawy, R., Stuckler, D., Mehta, A., Al-Ahdal, T., & Doebbeling, B. N. (2022). Effectiveness of early warning systems in the detection of infectious diseases outbreaks: A systematic review. *BMC Public Health*, 22(1), 2216. <https://doi.org/10.1186/s12889-022-14625-4>
- Melo, C. L., Mageste, L. R., Guaraldo, L., Paula, D. P., & Wakimoto, M. D. (2024). Use of Digital Tools in Arbovirus Surveillance: Scoping Review. *Journal of Medical Internet Research*, 26, e57476. <https://doi.org/10.2196/57476>

- Munn, Z., Barker, T. H., Moola, S., Tufanaru, C., Stern, C., McArthur, A., Stephenson, M., & Aromataris, E. (2020). Methodological quality of case series studies: An introduction to the JBI critical appraisal tool. *JBI Evidence Synthesis*, *18*(10), 2127-2133.
<https://doi.org/10.11124/JBISRIR-D-19-00099>
- Nageshwaran, G., Harris, R. C., & Guerche-Seblain, C. E. (2021). Review of the role of big data and digital technologies in controlling COVID-19 in Asia: Public health interest vs. privacy. *DIGITAL HEALTH*, *7*, 20552076211002953.
<https://doi.org/10.1177/20552076211002953>
- O'Shea, J. (2017). Digital disease detection: A systematic review of event-based internet biosurveillance systems. *International Journal of Medical Informatics*, *101*, 15-22.
- Perlaza, C. L., Cruz Mosquera, F. E., Moreno Reyes, S. P., Tovar Salazar, S. M., Cruz Rojas, A. F., España Serna, J. D., & Liscano, Y. (2024). Sociodemographic, Clinical, and Ventilatory Factors Influencing COVID-19 Mortality in the ICU of a Hospital in Colombia. *Healthcare*, *12*(22), 2294. <https://doi.org/10.3390/healthcare12222294>
- Poirel, L., Vuillemin, X., Kieffer, N., Mueller, L., Descombes, M.-C., & Nordmann, P. (2019). Identification of FosA8, a Plasmid-Encoded Fosfomycin Resistance Determinant from *Escherichia coli*, and Its Origin in *Leclercia adecarboxylata*. *ANTIMICROBIAL AGENTS AND CHEMOTHERAPY*, *63*(11). <https://doi.org/10.1128/AAC.01403-19>
- Ragnedda, M., Ruiu, M. L., & Calderón-Gómez, D. (2024). Examining the Interplay of Sociodemographic and Sociotechnical Factors on Users' Perceived Digital Skills. *Media and Communication*, *12*, 8167. <https://doi.org/10.17645/mac.8167>
- Salathé, M., Bengtsson, L., Bodnar, T. J., Brewer, D. D., Brownstein, J. S., Buckee, C., Campbell, E. M., Cattuto, C., Khandelwal, S., Mabry, P. L., & Vespignani, A. (2012).

- Digital epidemiology. *PLoS Computational Biology*, 8(7), e1002616.
<https://doi.org/10.1371/journal.pcbi.1002616>
- Samaras, L., García-Barriocanal, E., & Sicilia, M.-A. (2020). Comparing Social media and Google to detect and predict severe epidemics. *Scientific Reports*, 10(1), 4747.
<https://doi.org/10.1038/s41598-020-61686-9>
- Seo Dong-Woo, Shin Soo-Yong. (2017). Methods Using Social Media and Search Queries to Predict Infectious Disease Outbreaks. *Health Inform Res*, 23(4), 343-348.
<https://doi.org/10.4258/hir.2017.23.4.343>
- Shakeri Hossein Abad, Z., Kline, A., Sultana, M., Noaen, M., Nurmambetova, E., Lucini, F., Al-Jefri, M., & Lee, J. (2021). Digital public health surveillance: A systematic scoping review. *Npj Digital Medicine*, 4(1), 41. <https://doi.org/10.1038/s41746-021-00407-6>
- Sharpe, J. D., Hopkins, R. S., Cook, R. L., & Striley, C. W. (2016). Evaluating Google, Twitter, and Wikipedia as Tools for Influenza Surveillance Using Bayesian Change Point Analysis: A Comparative Analysis. *JMIR Public Health and Surveillance*, 2(2), e161.
<https://doi.org/10.2196/publichealth.5901>
- Shausan, A., Nazarathy, Y., & Dyda, A. (2023a). Emerging data inputs for infectious diseases surveillance and decision making. *Frontiers in Digital Health*, 5, 1131731.
<https://doi.org/10.3389/fdgth.2023.1131731>
- Shausan, A., Nazarathy, Y., & Dyda, A. (2023b). Emerging data inputs for infectious diseases surveillance and decision making. *Frontiers in Digital Health*, 5.
<https://doi.org/10.3389/fdgth.2023.1131731>
- Shaweno, D., Karmakar, M., Alene, K. A., Ragonnet, R., Clements, A. C., Trauer, J. M., Denholm, J. T., & McBryde, E. S. (2018). Methods used in the spatial analysis of

- tuberculosis epidemiology: A systematic review. *BMC Medicine*, 16(1), 193.
<https://doi.org/10.1186/s12916-018-1178-4>
- Shen, C., Chen, A., Luo, C., Zhang, J., Feng, B., & Liao, W. (2020). Using Reports of Symptoms and Diagnoses on Social Media to Predict COVID-19 Case Counts in Mainland China: Observational Inveillance Study. *Journal of Medical Internet Research*, 22(5), e19421.
<https://doi.org/10.2196/19421>
- Shin, S.-Y., Seo, D.-W., An, J., Kwak, H., Kim, S.-H., Gwack, J., & Jo, M.-W. (2016). High correlation of Middle East respiratory syndrome spread with Google search and Twitter trends in Korea. *Scientific Reports*, 6(1), 32920. <https://doi.org/10.1038/srep32920>
- Sulaiman, F., Yanti, N. S., Lesmanawati, D. A. S., Trent, M. J., Macintyre, C. R., & Chughtai, A. A. (2019). Language Specific Gaps in Identifying Early Epidemic Signals – A Case Study of the Malay Language. *Global Biosecurity*.
<https://api.semanticscholar.org/CorpusID:208083242>
- Syrowatka, A. (2021). *Leveraging artificial intelligence for pandemic preparedness and response: A scoping review to identify key use cases*.
- Tarkoma, S., Alghnam, S., & Howell, M. D. (2020). Fighting pandemics with digital epidemiology. *EClinicalMedicine*, 26, 100512.
<https://doi.org/10.1016/j.eclinm.2020.100512>
- Terry, K., Yang, F., Yao, Q., & Liu, C. (2023). The role of social media in public health crises caused by infectious disease: A scoping review. *BMJ Global Health*, 8(12).
<https://doi.org/10.1136/bmjgh-2023-013515>
- Timpka, T., Spreco, A., Dahlström, Ö., Eriksson, O., Gursky, E., Ekberg, J., Blomqvist, E., Strömberg, M., Karlsson, D., Eriksson, H., Nyce, J., Hinkula, J., & Holm, E. (2014).

- Performance of eHealth Data Sources in Local Influenza Surveillance: A 5-Year Open Cohort Study. *Journal of Medical Internet Research*, 16(4), e116.
<https://doi.org/10.2196/jmir.3099>
- Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D. J., Horsley, T., Weeks, L., Hempel, S., Akl, E. A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M. G., Garritty, C., ... Straus, S. E. (2018). PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Annals of Internal Medicine*, 169(7), 467-473. <https://doi.org/10.7326/M18-0850>
- Villanueva Parra, I., Muñoz Diaz, V., Martínez Guevara, D., Cruz Mosquera, F. E., Prieto-Alvarado, D. E., & Liscano, Y. (2024). A Scoping Review of Angiostrongyliasis and Other Diseases Associated with Terrestrial Mollusks, Including *Lissachatina fulica*: An Overview of Case Reports and Series. *Pathogens*, 13(10), 862.
<https://doi.org/10.3390/pathogens13100862>
- Yan, S. J., Chughtai, A. A., & Macintyre, C. R. (2017). Utility and potential of rapid epidemic intelligence from internet-based sources. *International Journal of Infectious Diseases*, 63, 77-87. <https://doi.org/10.1016/j.ijid.2017.07.020>
- Yousefinaghani, S., Dara, R., Mubareka, S., & Sharif, S. (2021). Prediction of COVID-19 Waves Using Social Media and Google Search: A Case Study of the US and Canada. *Frontiers in Public Health*, 9, 656635. <https://doi.org/10.3389/fpubh.2021.656635>
- Zhao, I. Y., Ma, Y. X., Yu, M. W. C., Liu, J., Dong, W. N., Pang, Q., Lu, X. Q., Molassiotis, A., Holroyd, E., & Wong, C. W. W. (2021). Ethics, Integrity, and Retributions of Digital

Detection Surveillance Systems for Infectious Diseases: Systematic Literature Review.

Journal of Medical Internet Research, 23(10), e32328. <https://doi.org/10.2196/32328>