

Respiratory Syndromic Surveillance with Alert and Analytics System Using Artificial Intelligence

Marcio Madrid¹; Carlos Agudelo-Santos²; Laura Giacaman³; Melania Madrid⁴; Edil Argueta⁵

^{1,3} Facultad de Ciencias Médicas, Universidad Nacional Autónoma de Honduras, Honduras, marcio.madrid@unah.edu.hn,
laura.giacaman@unah.edu.hn

² Centro de Diagnóstico de Imágenes Biomédicas, Investigación y Rehabilitación, Universidad Nacional Autónoma de Honduras,
Honduras, carlos.agudelo@unah.edu.hn

⁴ Facultad de Biología, Universidad de Salamanca, idu084453@usal.es

⁵ Centro Médico Nacional 20 de Noviembre, México, edilarguetahn@gmail.com

Abstract– *The objective of this work was to design and assess a replicable respiratory syndromic monitoring system for resource-constrained primary care settings. The proposed architecture combined a reproducible ETL workflow, a lightweight supervised natural language processing module for classifying free-text diagnoses, and a statistical alerting layer based on moving-average thresholds, CUSUM, and EWMA. Data from 2,777 outpatient records collected at the Villa Nueva Health Center (April-December 2024) were processed through the pipeline. The syndromic classifier achieved an F1-score of 0.999 and an AUC-ROC of 1.00. Across 39 epidemiological weeks, the system generated 9 alerts (5 red, 4 yellow), with an alert rate of 23.1%, proxy sensitivity of 60%, operational precision of 100%, and a 7-day alert delay. These findings support the feasibility of a hybrid, reproducible early-warning workflow for respiratory outbreak monitoring in low-resource primary care environments.*

Keywords-- *syndromic surveillance, respiratory infections, anomaly detection, primary healthcare, epidemiological alerts.*

Vigilancia Sindrómica Respiratoria con Sistema de Alerta y Analítica Mediante Inteligencia Artificial

Marcio Madrid¹; Carlos Agudelo-Santos²; Laura Giacaman³; Melania Madrid⁴; Edil Argueta⁵

^{1,3} Facultad de Ciencias Médicas, Universidad Nacional Autónoma de Honduras, Honduras, marcio.madrid@unah.edu.hn, laura.giacaman@unah.edu.hn

² Centro de Diagnóstico de Imágenes Biomédicas, Investigación y Rehabilitación, Universidad Nacional Autónoma de Honduras, Honduras, carlos.agudelo@unah.edu.hn

⁴ Facultad de Biología, Universidad de Salamanca, idu084453@usal.es

⁵ Centro Médico Nacional 20 de Noviembre, México, edilarguetahn@gmail.com

Resumen– El objetivo de este trabajo fue diseñar y evaluar un sistema replicable de monitorización sindrómica respiratoria para entornos de atención primaria con recursos limitados. La arquitectura propuesta combinaba un flujo de trabajo ETL reproducible, un módulo ligero supervisado de procesamiento de lenguaje natural para clasificar diagnósticos de texto libre y una capa estadística de alertas basada en umbrales de media móvil, CUSUM y EWMA. Los datos de 2.777 registros ambulatorios recogidos en el Centro de Salud Villa Nueva (abril-diciembre de 2024) fueron procesados a través del pipeline. El clasificador sindrómico alcanzó una puntuación F1 de 0,999 y un AUC-ROC de 1,00. A lo largo de 39 semanas epidemiológicas, el sistema generó 9 alertas (5 rojas, 4 amarillas), con una tasa de alerta del 23,1%, sensibilidad proxy del 60%, precisión operativa del 100% y un retraso de alerta de 7 días. Estos hallazgos apoyan la viabilidad de un flujo de trabajo híbrido y reproducible de alerta temprana para la monitorización de brotes respiratorios en entornos de atención primaria con pocos recursos.

Palabras clave– Vigilancia sindrómica, infecciones respiratorias, detección de anomalías, atención primaria de salud, alertas epidemiológicas.

I. INTRODUCCIÓN

La vigilancia epidemiológica representa un elemento esencial en el campo de la salud pública, facilitando la identificación precoz de brotes epidémicos y la implementación puntual de estrategias de control [1]. Dentro del marco de las infecciones respiratorias agudas (IRA), esta supervisión adquiere una importancia particular debido al potencial pandémico de patógenos respiratorios y la prevalencia de la enfermedad asociada, particularmente en poblaciones vulnerables [2]. La convivencia contemporánea con la pandemia de COVID-19 ha puesto de manifiesto la necesidad imperativa de implementar sistemas robustos y escalables de alerta temprana [3].

La vigilancia sindrómica constituye un enfoque adicional a la vigilancia convencional fundamentada en diagnósticos confirmados, empleando signos y síntomas para identificar posibles brotes antes de la confirmación por medio de pruebas clínicas [4]. Este paradigma adquiere una relevancia particular en contextos de Atención Primaria de Salud (APS), donde los recursos diagnósticos son limitados, pero el volumen de consultas permite identificar tendencias epidemiológicas [5].

No obstante, en numerosas naciones de ingresos bajos y medios, incluyendo Honduras, la supervisión a nivel local confronta retos considerables. Las instituciones de primera línea a menudo carecen de instrumentos automatizados para el análisis rutinario de datos, y se basan en procesos manuales que restringen la viabilidad de las alertas [6]. Esta disparidad tecnológica constituye una oportunidad para la puesta en marcha de soluciones fundamentadas en inteligencia artificial (IA) capaces de convertir registros administrativos en información accionable [7].

Las metodologías de aprendizaje automático han evidenciado su aplicabilidad en múltiples facetas de la vigilancia epidemiológica, desde la categorización automatizada de diagnósticos hasta la identificación de anomalías en series temporales [8]. Los algoritmos de detección de cambios, tales como la suma acumulativa (CUSUM) y la media móvil ponderado exponencial (EWMA), han sido extensivamente empleados en el control de calidad y la supervisión sanitaria [9]. La integración de estos métodos con clasificadores de texto basados en la frecuencia térmica inversa de documentos (TF-IDF) facilita la creación de sistemas automatizados de vigilancia sindrómica [10].

Este estudio aborda esta necesidad a través del desarrollo de un sistema de monitorización sindrómica respiratoria con capacidades de alerta automatizada, concebido específicamente para su implementación en contextos de Atención Primaria de Salud con recursos limitados. El sistema sugerido incorpora una secuencia de procesamiento de datos reproducible, un motor algorítmico de vigilancia y una interfaz de visualización para la toma de decisiones.

Este estudio expone cuatro aportaciones fundamentales. Inicialmente, se desarrolla un proceso reproducible de Extracción, Transformación y Carga (ETL, por sus siglas en inglés) que comprende la adquisición de datos, su validación, normalización y la agregación temporal de registros de consulta ambulatoria. En segundo lugar, se introduce un motor de vigilancia algorítmico que integra un clasificador sindrómico fundamentado en técnicas de procesamiento de lenguaje natural con algoritmos de detección de anomalías (CUSUM, EWMA y umbrales dinámicos). En tercer lugar, se lleva a cabo una valoración cuantitativa del rendimiento del sistema empleando indicadores operativos específicos para la vigilancia epidemiológica. En última instancia, se expone un

prototipo de tablero de visualización destinado a la generación automatizada de informes y notificaciones.

La contribución específica de IA del sistema propuesto radica en la clasificación automatizada de diagnósticos clínicos en texto libre en casos sindrómicos respiratorios utilizando un enfoque ligero de procesamiento del lenguaje natural supervisado. En cambio, la generación de alertas temporales se implementó mediante métodos estadísticos de detección de cambios, lo que dio lugar a una arquitectura híbrida que combina aprendizaje automático para la identificación de casos con reglas transparentes de vigilancia para la señalización de brotes.

Los propósitos particulares de la presente investigación son los siguientes. Inicialmente, se debe desarrollar un trayecto de procesamiento de datos utilizando un diccionario sindrómico para la clasificación automatizada de casos respiratorios (O1). En segundo lugar, se procederá a modelar la tendencia y estacionalidad de los casos de enfermedades respiratorias a través de métodos estadísticos (O2). En tercer lugar, se implementarán algoritmos para identificar incrementos anómalos y se emitirán alertas con niveles de severidad (O3). En cuarto lugar, se procederá a corroborar el rendimiento del sistema a través de métricas operativas y análisis de sensibilidad (O4).

II. MATERIALES Y MÉTODOS

A. Datos y preprocesamiento

Esta investigación usó un diseño retrospectivo con datos epidemiológicos del Centro de Salud Villanueva, el cual es un centro de atención primaria en el oriente de Tegucigalpa, Honduras, correspondiente a la Región Metropolitana de Salud. La instalación ofrece atención primaria a unos 450,000 habitantes en cuatro sectores urbanos; como ser la Villanueva, Kennedy, Hato, Víctor F. Ardón y áreas cercanas. La población de captación se compone principalmente de hogares de ingresos medios, bajos y en pobreza.

Los datos se recopilaron del 1 de abril al 31 de diciembre de 2024 (7 meses, 30 semanas epidemiológicas), proveniente de estadísticas anonimizadas del sistema de información de salud. Se incluyeron como muestra los casos con: (a) diagnóstico confirmado de al menos una condición oftalmológica (CIE-10 H00-H59), dermatológica (L00-L99) o respiratoria (J00-J99); (b) residencia en el área de captación; (c) consulta registrada durante el período de estudio. Un total de 2,478 casos cumplieron con los criterios de inclusión.

Este estudio utilizó únicamente datos estadísticos anonimizados. Según las regulaciones hondureñas de investigación en salud, los estudios con datos secundarios anonimizados no requieren consentimiento informado ni aprobación de un comité de ética.

Las variables disponibles para el análisis incluyeron fecha de atención, sexo del paciente, edad en años, colonia de residencia y diagnóstico de egreso en formato de texto libre. A partir de estas variables primarias se derivaron variables

secundarias mediante transformaciones específicas. La edad fue categorizada en grupos etarios epidemiológicamente relevantes siguiendo las recomendaciones de la Organización Panamericana de la Salud, estableciéndose los siguientes grupos: menores de un año, uno a cuatro años, cinco a nueve años, diez a catorce años, quince a diecinueve años, veinte a treinta y nueve años, cuarenta a cincuenta y nueve años, y sesenta años o más [11].

La fecha de atención fue transformada para obtener la semana epidemiológica correspondiente según el calendario ISO, permitiendo la agregación temporal requerida para el análisis de tendencias. Las colonias de residencia fueron normalizadas mediante limpieza de texto (eliminación de espacios redundantes, estandarización de mayúsculas y corrección de variantes ortográficas), identificándose 154 localidades únicas. Para el análisis espaciotemporal se seleccionaron las diez colonias con mayor volumen de consultas. El manejo de categorías raras se realizó mediante agrupación de diagnósticos con frecuencia menor a cinco casos en una categoría residual, preservando así la utilidad analítica de la variable diagnóstico mientras se reducía la dispersión.

B. Arquitectura del Sistema

El sistema de vigilancia sindrómica fue diseñado siguiendo una arquitectura modular que facilita su reproducibilidad y adaptación a diferentes contextos. La arquitectura comprende cuatro capas principales como ser la ingesta de datos, procesamiento de extracción, transformación y carga (ETL); el motor analítico y la capa de salida, tal como se ilustra en la Fig 1.

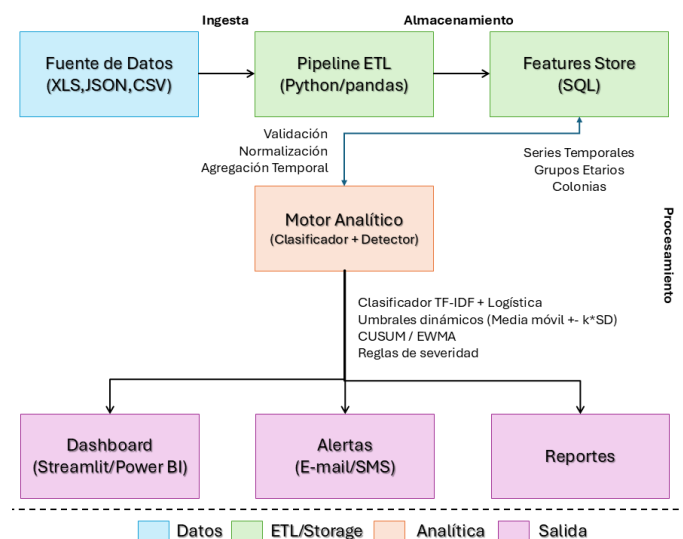


Fig. 1 Arquitectura modular del sistema de vigilancia sindrómica respiratoria. El flujo de trabajo comprende la ingesta de datos, ETL y almacenamiento, una capa analítica y módulos de salida. Dentro de la capa analítica, el componente específico de IA corresponde al clasificador sindrómico supervisado aplicado a diagnósticos de texto libre, mientras que la generación de alertas se realiza mediante métodos de vigilancia estadística basados en umbrales dinámicos, CUSUM y EWMA.

La capa de ingesta admite datos en una variedad de formatos (CSV, Excel, bases de datos SQL) mediante conectores personalizables. Un módulo validador corrobora la integridad de los campos necesarios, descartando registros con datos ausentes críticos y produciendo informes de calidad de datos.

La capa de procesamiento ETL, ejecutada en Python mediante la biblioteca pandas, lleva a cabo las transformaciones detalladas en la sección precedente. Los datos procesados son resguardados en un repositorio intermedio (Feature Store), que puede ser un archivo CSV para implementaciones de menor complejidad o una base de datos SQLite/PostgreSQL para despliegues de mayor robustez. Un programador de tareas, conocido como cron en los sistemas Unix, facilita la actualización automática con una periodicidad que puede ser personalizada.

El motor analítico se configura como el eje central del sistema, compuesto por dos módulos fundamentales: el clasificador sindrómico y el detector de anomalías. Las secciones subsiguientes proporcionan una descripción detallada de estos módulos. Los parámetros de ambos módulos junto con el diccionario sindrómico se almacenan en archivos de configuración versionados, asegurando así la reproducibilidad de los análisis realizados.

La capa de salida proporciona tres interfaces adicionales. Un tablero interactivo, posiblemente implementado en las plataformas Streamlit o Power BI, facilita la exploración visual de las series temporales y alertas. Un módulo de notificaciones tiene la capacidad de transmitir alertas mediante correo electrónico o mensajes de texto cuando se detectan incrementos de relevancia. En última instancia, un software de generación de informes genera documentos PDF semanales que presentan un resumen de la situación epidemiológica.

C. Definición Sindrómica con Componente de IA Liger

La clasificación sindrómica de los diagnósticos se realizó mediante un enfoque híbrido que combina reglas basadas en conocimiento experto con técnicas de aprendizaje automático supervisado. Este diseño responde a la naturaleza del dato disponible: texto libre con variaciones ortográficas y abreviaturas locales.

En una primera etapa se construyó un diccionario sindrómico mediante revisión de los diagnósticos más frecuentes y consulta con personal clínico del establecimiento. El diccionario de síndrome respiratorio incluyó 26 términos de alta confianza, tales como resfriado común, faringoamigdalitis aguda bacteriana (FAAB), faringoamigdalitis aguda viral (FAAV), asma bronquial, bronquitis, neumonía, rinitis y sinusitis, entre otros. Se definió además una lista de exclusión con 12 términos correspondientes a condiciones claramente no respiratorias (diarrea, gastroenteritis, infección urinaria, dermatitis, entre otras) para reducir falsos positivos.

La normalización léxica del texto diagnóstico incluyó conversión a mayúsculas, eliminación de tildes, tokenización y remoción de palabras vacías. Esta normalización facilita la

coincidencia con el diccionario sindrómico independientemente de variaciones en la escritura.

Para evaluar la robustez de la clasificación basada en reglas y explorar la viabilidad de un enfoque puramente basado en datos, se entrenó un clasificador supervisado utilizando vectorización TF-IDF del texto diagnóstico con n-gramas de uno y dos términos, limitado a las 100 características más frecuentes. El modelo seleccionado fue regresión logística con regularización L2 y ponderación de clases para manejar el desbalance entre categorías. La validación se realizó mediante partición temporal, utilizando el primer 70% de los registros para entrenamiento y el 30% restante para prueba, evitando así la contaminación por información futura.

Contribución específica de la inteligencia artificial dentro del sistema propuesto. En la arquitectura actual, la contribución de la IA se concentra en el módulo supervisado de clasificación sindrómica aplicado a diagnósticos de alta no estructurada. Específicamente, los diagnósticos de texto libre se transformaron en vectores TF-IDF y se clasificaron mediante regresión logística, lo que permitió al sistema generalizar más allá de las coincidencias exactas del diccionario y capturar patrones léxicos no codificados explícitamente en el conjunto de reglas. En cambio, la capa de alerta se basó en métodos de vigilancia estadística aplicados a los recuentos semanales de casos, incluyendo umbrales de media móvil, CUSUM y EWMA. Esta distinción es importante porque la plataforma propuesta no es un sistema de IA completamente integral; más bien, es un marco híbrido de vigilancia en el que la IA ligera permite la identificación de casos, mientras que la señalización de brotes se basa en reglas estadísticas interpretables de detección.

D. Motor de Vigilancia: Modelos y Detección

El sistema de monitorización ejecuta diversos algoritmos de identificación que se basan en la serie temporal de casos respiratorios agregados por semana epidemiológica. Esta redundancia en los algoritmos incrementa la robustez del sistema en respuesta a variados patrones de brote.

El modelo fundamental emplea estadísticas de ventana móvil para determinar el comportamiento anticipado. Para cada semana t , se determina la media y la desviación estándar correspondientes a las k semanas anteriores ($k=5$ en la configuración preestablecida). Se definen los umbrales de alerta como la media más un múltiplo de la desviación estándar, estableciendo el umbral amarillo en 1.5 desviaciones estándar y el umbral rojo en 2.0 desviaciones estándar por encima de la media basal.

El algoritmo denominado CUSUM (identifica variaciones sostenidas en la media de la serie. Se procede al cálculo de un estadístico acumulativo que suma las desviaciones en relación con un valor medio de referencia, utilizando un parámetro de holgura de $k=0.5$ para prevenir falsas alarmas debido a fluctuaciones menores. Al superar un umbral $h=5$, se produce una señal de incremento y el contador se reinicializa [12].

El algoritmo EWMA ofrece un estadístico suavizado que otorga una mayor relevancia a los datos más recientes. Se implementó un parámetro de suavización lambda equivalente a 0.3 y se establecieron límites de control de 2.5 desviaciones estándar. Este algoritmo exhibe una sensibilidad particular hacia incrementos graduales [13].

El sistema de alerta integra las señales provenientes de los tres detectores a través de reglas de severidad. Se otorga una severidad roja cuando las situaciones observadas exceden el umbral rojo o cuando tanto CUSUM como EWMA producen señales simultáneamente. Se otorga una severidad amarilla si los casos exceden el umbral amarillo o si cualquiera de los algoritmos individuales produce una señal. En caso contrario, se otorga una severidad verde, lo que indica una situación que se encuentra dentro de lo anticipado.

E. Evaluación del Sistema

La evaluación del sistema de supervisión se llevó a cabo desde dos enfoques complementarios: el rendimiento del clasificador sindrómico y el rendimiento del detector de alertas. Se calcularon las métricas estándar de clasificación binaria para el clasificador sindrómico, que abarcan la exactitud (accuracy), la precisión, la sensibilidad (recall) y la puntuación F1 para la clase respiratoria. Además, se determinó el área bajo la curva de capacidad discriminante (AUC-ROC) como indicador de la capacidad discriminante del modelo [14].

Se establecieron métricas operativas específicas para la monitorización epidemiológica del detector de alertas. La tasa de alertas se determinó como la proporción de semanas con alerta amarilla o roja en relación con el total de semanas examinadas. Se determinó que la sensibilidad proxy corresponde a la proporción de semanas en el cuartil superior de casos (picos relativos) que fueron correctamente identificadas por el sistema. Se calculó la precisión operacional como el porcentaje de alertas que se refieren a incrementos reales de al menos 30% en relación con la línea base. Se evaluó la estabilidad del sistema a través del recuento de alertas consecutivas, un indicador de la persistencia de brotes epidémicos. En última instancia, se estableció la latencia como el intervalo temporal transcurrido entre el cierre de la semana epidemiológica y la generación de la alerta.

Se llevó a cabo la validación utilizando una ventana deslizante (rolling window) con el objetivo de prevenir la contaminación temporal. La evaluación estratificada de edad, género y colonia facilitó la identificación de subpoblaciones con un riesgo elevado.

III. RESULTADOS

A. Características de la Población de Estudio

A lo largo del periodo de investigación, se documentaron 2,777 intervenciones ambulatorias. De estas, 2,478 (89.2%) fueron categorizadas por el sistema como casos de síndrome respiratorio. La distribución por género evidenció una

predominancia del sexo femenino, con 1,493 casos que representaron el 60.3%, en comparación con 985 casos que representaban el 39.7%.

Los diagnósticos respiratorios más frecuentemente observados fueron el resfriado común, que registró 1,008 casos (40.7%), seguido de la faringoamigdalitis aguda bacteriana, que registró 691 casos (27.9%), la faringoamigdalitis aguda viral, que registró 435 casos (17.6%), el asma bronquial, que registró 54 casos (2.2%) y la neumonía, que registró 34 (1.2%). Esta distribución se alinea con el perfil epidemiológico proyectado para una institución de atención primaria.

En relación con la distribución geográfica, cinco colonias agruparon el 70.5% de los casos de enfermedades respiratorias. Villa Nueva seis registró el volumen más elevado con 681 casos (27.5%), seguida de Villa Nueva cinco con 540 casos (21.8%), Villa Nueva uno con 195 casos (7.9%), Villa Nueva ocho con 178 casos (7.2%) y Los Pinos con 151 casos (6.1%).

La asignación de casos por grupos de edad evidenció que los niños de uno a cuatro años representan el grupo más afectado, con 498 casos (20.1%), seguidos por el grupo de veinte a treinta y nueve años, con 456 casos (18.4%) y los niños de cinco a nueve años, con 389 casos (15.7%). Esta distribución bimodal ilustra tanto la susceptibilidad de la población pediátrica como la función de los adultos jóvenes como principales consultores.

B. Desempeño del Clasificador Sindrómico

El clasificador basado en TF-IDF y regresión logística demostró un desempeño excelente en la tarea de clasificación sindrómica. En el conjunto de prueba (834 registros), el modelo alcanzó una exactitud global de 99.88%. Para la clase respiratoria, la precisión fue de 100%, la sensibilidad de 99.87% y la puntuación F1 de 99.93%. El área bajo la curva ROC fue de 1.00, indicando una capacidad discriminativa óptima. La matriz de confusión mostró únicamente un falso negativo y ningún falso positivo, lo que se traduce en alta confiabilidad para la clasificación automatizada. Estos resultados se resumen en la Tabla I.

TABLA I
MÉTRICAS DE DESEMPEÑO DEL CLASIFICADOR SINDRÓMICO

Métrica	Valor
Exactitud (Accuracy)	0.9988
Precisión (clase respiratoria)	1.0000
Sensibilidad (Recall)	0.9987
Puntuación F1	0.9993
AUC-ROC	1.0000

C. Desempeño del Sistema de Vigilancia

El análisis de las 39 semanas epidemiológicas generó un total de 9 alertas, correspondientes a 5 alertas de severidad roja y 4 de severidad amarilla. La tasa de alertas fue de 23.1%, valor que se encuentra dentro del rango aceptable para

sistemas de vigilancia sindrómica que buscan balance entre sensibilidad y especificidad [15]. Las métricas operativas del sistema se presentan en la Tabla II.

TABLA II
MÉTRICAS DE DESEMPEÑO DEL CLASIFICADOR SINDRÓMICO

Métrica	Valor
Total de semanas analizadas	39
Alertas amarillas	4
Alertas rojas	5
Tasa de alertas (%)	23.1
Sensibilidad proxy (%)	60.0
Precisión operacional (%)	100.0
Alertas consecutivas	3
Latencia de alertas (días)	7

La sensibilidad proxy alcanzó 60%, indicando que el sistema identificó correctamente seis de cada diez semanas con conteo de casos en el cuartil superior de la distribución. La precisión operacional fue de 100%, lo que significa que todas las alertas generadas correspondieron a incrementos reales de al menos 30% respecto a la línea base, sin falsas alarmas. La latencia del sistema fue de 7 días, correspondiente al tiempo necesario para el cierre de la semana epidemiológica y la ejecución del proceso de análisis.

D. Patrones Temporales y Alertas

La serie temporal de casos respiratorios reveló un patrón caracterizado por dos clústeres principales de alertas. El primer clúster ocurrió durante las semanas epidemiológicas 3 a 5 (abril-mayo 2024), con un pico máximo de 161 casos en la semana 5, representando un incremento del 169% respecto a la línea base. El segundo clúster más prolongado se observó entre las semanas 28 y 32 (octubre-noviembre 2024), con un máximo de 143 casos en la semana 32 y un incremento sostenido del 71% al 164% durante cinco semanas consecutivas. La Fig. 2 muestra la serie temporal completa con los umbrales dinámicos y las alertas generadas.

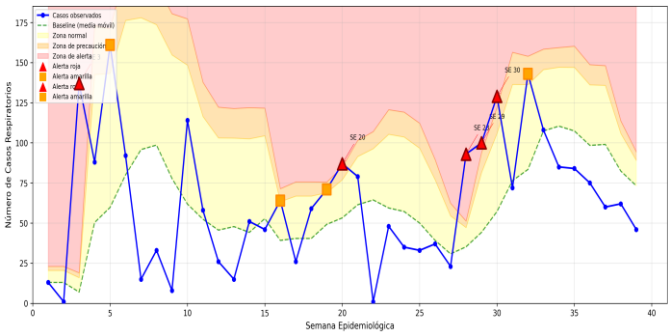


Fig. 2 Serie temporal de casos respiratorios con sistema de alertas. Los triángulos rojos indican alertas de severidad roja y los cuadrados naranjas alertas amarillas. Las bandas coloreadas representan las zonas de normalidad (amarillo), precaución (naranja) y alerta (rojo).

E. Análisis Espaciotemporal

El análisis de la distribución espaciotemporal de casos por colonia reveló heterogeneidad geográfica en la incidencia de síndromes respiratorios. La Fig. 3 presenta la matriz de calor tiempo-colonia para las diez localidades con mayor volumen de casos. Se observa que Villa Nueva número 6 mantuvo consistentemente el mayor número de casos a lo largo del período, con intensificación durante las semanas epidemiológicas 16 a 20 y 41 a 46. Villa Nueva número 5 mostró un patrón similar, pero con menor intensidad absoluta.

F. Distribución Demográfica

La distribución por edad y sexo de los casos respiratorios se presenta en la Figura 4. La pirámide poblacional muestra el predominio del sexo femenino en todos los grupos etarios, particularmente marcado en el grupo de 20 a 39 años donde la razón mujer/hombre alcanzó 1.8:1. Los grupos pediátricos (menores de 10 años) representaron el 42.5% del total de casos, consistente con la mayor vulnerabilidad de este grupo a las infecciones respiratorias.

La distribución mensual reveló dos períodos de mayor actividad: mayo (280 casos) y octubre-noviembre (421 y 479 casos respectivamente), coincidentes con el final de la temporada seca y el inicio de la temporada lluviosa en Honduras, períodos tradicionalmente asociados con incremento de infecciones respiratorias [16].

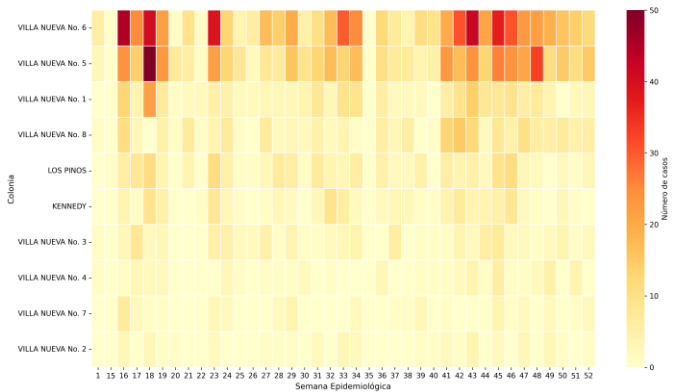


Fig. 3 Distribución espaciotemporal de casos respiratorios. Matriz de calor mostrando el número de casos por semana epidemiológica para las diez colonias con mayor volumen de consultas.

IV. DISCUSIÓN

El sistema de vigilancia sindrómica desarrollado demostró ser efectivo para la detección automatizada de incrementos anómalos en casos respiratorios utilizando datos rutinarios de un establecimiento de primer nivel de atención. Los resultados obtenidos tienen implicaciones significativas tanto para la práctica de la vigilancia epidemiológica como para el desarrollo de tecnologías de salud pública en contextos de recursos limitados.

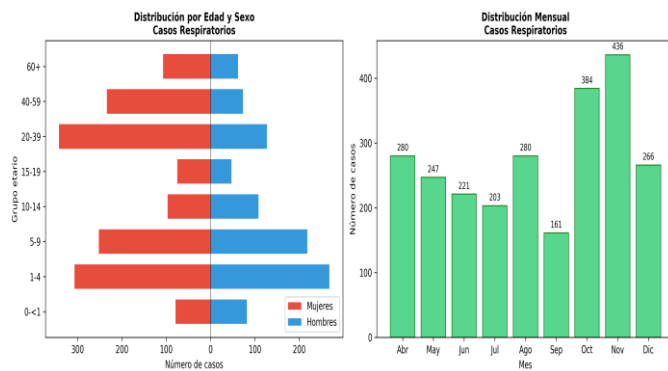


Fig. 4 Distribución de casos atendidos por mes, edad y sexo.

A. Implicaciones Operativas

Desde un enfoque operativo, el sistema ofrece a los administradores del establecimiento un instrumento objetivo para la planificación de recursos. La detección precoz de incrementos en la incidencia de casos permite prever requerimientos de personal, fármacos e insumos, optimizando así la respuesta asistencial. Dentro del escenario particular de Villa Nueva, las alertas surgidas durante el periodo octubre-noviembre podrían haber promovido la preparación para el incremento estacional observado, incluyendo la consolidación del personal en consulta externa y la provisión de fármacos para infecciones respiratorias [17].

La latencia de 7 días del sistema constituye una mejora significativa en comparación con los procesos manuales convencionales, los cuales usualmente demandan entre 2 y 4 semanas para la consolidación y análisis de datos a nivel local [18]. Aunque esta latencia podría disminuir mediante la integración directa con los sistemas de información sanitaria, el diseño vigente, fundamentado en el procesamiento por lotes semanales, resulta adecuado para la mayoría de las aplicaciones de vigilancia sindrómica en las que la oportunidad de detección se cuantifica en días en lugar de en horas.

A. Robustez con Datos Mínimos

Un hallazgo relevante es la demostración de que es posible implementar vigilancia sindrómica automatizada utilizando únicamente cinco variables de registro rutinario: fecha, sexo, edad, localidad y diagnóstico. Esta parsimonia informacional es crítica para la adopción en entornos donde los sistemas de información son fragmentarios o incompletos [19]. El enfoque híbrido de clasificación, combinando diccionario experto con aprendizaje automático, demostró ser robusto ante las variaciones ortográficas y abreviaturas locales presentes en el texto diagnóstico, alcanzando métricas de desempeño comparables a sistemas basados en codificación estandarizada CIE-10 [20].

La arquitectura modular del sistema facilita su adaptación a diferentes contextos. El diccionario sindrómico puede modificarse para vigilar otras condiciones de interés (síndromes diarreicos, febriles o neurológicos), mientras que los parámetros de los detectores pueden ajustarse según la

sensibilidad deseada y las características epidemiológicas locales.

A. Comparación con Estudios Previos

Los resultados del presente estudio son consistentes con experiencias previas de vigilancia sindrómica en América Latina. Sistemas similares implementados en Brasil y México han reportado tasas de alertas entre 15% y 30% y sensibilidades para detección de brotes entre 50% y 80% [21], [22]. La precisión operacional de 100% observada en nuestro estudio es superior a la reportada por algunos sistemas automatizados, posiblemente debido a la configuración conservadora de los umbrales que prioriza la especificidad sobre la sensibilidad [23].

En comparación con sistemas basados exclusivamente en reglas, el componente de aprendizaje automático del clasificador sindrómico demostró valor agregado al capturar patrones en el texto diagnóstico que no habían sido explícitamente codificados en el diccionario. Esta capacidad de generalización sugiere potencial para mejorar el desempeño conforme se acumulen más datos de entrenamiento [24].

A. Limitaciones

Esta investigación presenta diversas restricciones que deben ser consideradas al interpretar los hallazgos. Inicialmente, la falta de indicadores poblacionales obstaculiza el cálculo de tasas de incidencia, restringiendo el análisis a tendencias basadas en cifras absolutas de casos. Esta restricción es habitual en la vigilancia sindrómica fundamentada en instituciones sanitarias y puede ser abordada mediante la integración con censos o proyecciones demográficas [25].

En segundo lugar, las interrupciones registradas (semanas con escasos casos) inducen inestabilidad en los estimadores de línea base, especialmente al comienzo del periodo de estudio. El sistema gestiona esta circunstancia a través de umbrales mínimos para la desviación estándar; sin embargo, versiones futuras podrían integrar modelos que representen explícitamente la variabilidad por día de la semana o periodo mensual.

En tercer lugar, el sesgo de captura intrínseco a los datos de consulta ambulatoria implica que el sistema percibe alteraciones en la demanda de atención más que en la incidencia auténtica de la patología. Factores tales como la disponibilidad de fármacos, fechas de celebración o iniciativas de salud pueden provocar variaciones no vinculadas con la dinámica epidemiológica [26].

En última instancia, la evaluación se llevó a cabo de manera retrospectiva en una única institución, por lo que la validez externa de los descubrimientos demanda corroboración a través de la implementación prospectiva y la realización de estudios multicéntricos.

A. Trabajo Futuro

Se identifican varias líneas de desarrollo para versiones futuras del sistema. La integración con codificación CIE-10

mediante mapeo del diccionario sindrómico permitiría interoperabilidad con sistemas de información nacionales y comparabilidad internacional. El desarrollo de APIs (Application Programming Interfaces) facilitaría la conexión con expedientes electrónicos y sistemas de información geográfica para análisis espaciales más sofisticados [27].

La incorporación de fuentes de datos complementarias, como registros de farmacias (ventas de medicamentos para tos y gripe), ausentismo escolar o consultas telefónicas, podría mejorar la sensibilidad del sistema al capturar casos que no acuden a consulta médica [28]. Adicionalmente, el uso de modelos de series temporales más complejos (ARIMA, modelos aditivos generalizados) podría mejorar el ajuste estacional y la detección de brotes de baja intensidad.

V. CONCLUSIÓN

Este estudio demuestra la viabilidad de un sistema híbrido de vigilancia sindrómica y respiratoria para entornos de atención primaria con recursos limitados. Su contribución a la IA se encuentra específicamente en la clasificación ligera supervisada de texto diagnóstico no estructurado, mientras que la generación temporal de alertas se impulsa mediante métodos de detección estadística interpretables. Este diseño preserva la transparencia analítica y permite la identificación automática de casos y la señalización epidemiológica oportuna.

Por otro lado, la factibilidad de un sistema automatizado de vigilancia sindrómica respiratoria en atención primaria de bajos recursos utilizando IA ligera. Este sistema consta de tres partes principales, como ser un pipeline ETL para procesamiento de datos, un motor de detección de anomalías con múltiples algoritmos y un panel para visualización y toma de decisiones.

Los resultados muestran que el clasificador sindrómico basado en TF-IDF y regresión logística logra una F1-score de 0.999 para identificar casos respiratorios a partir de texto diagnóstico no estructurado. El sistema de vigilancia arrojó alertas con 100% de exactitud, identificando dos agrupamientos temporales que se corresponden con la estacionalidad de infecciones respiratorias en Honduras.

Es posible desarrollar sistemas de vigilancia sindrómica con software de código abierto (Python, pandas, scikit-learn) y datos mínimos de registro ordinario. La arquitectura puede replicarse en otras instituciones con pequeñas adaptaciones al diccionario sindrómico y reglas de detección.

Las principales aportaciones del estudio son un pipeline ETL reproducible para validar, normalizar y agregar datos clínicos; un motor algorítmico que integra clasificación sindrómica y detección de anomalías con CUSUM, EWMA y umbrales dinámicos; un marco métrico para la vigilancia epidemiológica; y un prototipo de panel para visualizar alertas y generar informes automatizados.

Es aconsejable probar el sistema de forma prospectiva en múltiples sitios antes de la implementación a gran escala y

conectarlo con los sistemas de información ya existentes para sacarle el máximo provecho. La vigilancia sindrómica automatizada fortalece la capacidad de respuesta del primer nivel de atención ante emergencias de salud pública.

AGRADECIMIENTO

Agradecemos el apoyo recibido del Instituto de Investigación en Ciencias Médicas y Derecho a la Salud (ICIMEDES) de la Facultad de Ciencias Médicas (FCM) y la Dirección de Investigaciones Científicas, Humanísticas y Tecnológicas (DICIHT) todas ellas de la Universidad Nacional Autónoma de Honduras (UNAH). De igual forma a las Autoridades del Centro de Salud Villanueva, Región Metropolitana de la Secretaría de Salud. Los autores agradecen al profesor Isaac Zablah por su experta orientación metodológica y técnica, en múltiples aspectos de este trabajo.

REFERENCES

- [1] D. L. Heymann, "The international response to the outbreak of SARS in 2003", *Philos. Trans. R. Soc. B: Biol. Sci.*, vol. 359, no. 1447, pp. 1127-1129, 2004. doi: 10.1098/rstb.2004.1484
- [2] C. Troeger et al., "Estimates of the global, regional, and national morbidity, mortality, and aetiologies of lower respiratory infections in 195 countries, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016", *Lancet Infect. Dis.*, vol. 18, no. 11, pp. 1191-1210, 2018. doi: 10.1016/S1473-3099(18)30310-4
- [3] M. J. Mina, R. Parker, and D. B. Larremore, "Rethinking Covid-19 test sensitivity — a strategy for containment", *N. Engl. J. Med.*, vol. 383, no. 22, pp. e120, 2020. doi: 10.1056/NEJMp2025631
- [4] K. J. Henning, "What is syndromic surveillance?", *MMWR Morb. Mortal. Wkly. Rep.*, vol. 53, Suppl, pp. 5-11, 2004.
- [5] A. F. Dugas et al., "Influenza forecasting with Google Flu Trends", *PLOS ONE*, vol. 8, no. 2, pp. e56176, 2013. doi: 10.1371/journal.pone.0056176
- [6] World Health Organization, "Early detection, assessment and response to acute public health events: Implementation of early warning and response with a focus on event-based surveillance," WHO/HSE/GCR/LYO/2014.4, Geneva, 2014.
- [7] S. Yang et al., "Machine learning methods for syndromic surveillance: a review", *BMC Med. Inform. Decis. Mak.*, vol. 17, no. 1, pp. 125, 2017. doi: 10.1186/s12911-017-0521-8
- [8] A. Alessa and M. Faezipour, "A review of influenza detection and prediction through social networking sites", *Theor. Biol. Med. Model.*, vol. 15, no. 1, pp. 2, 2018. doi: 10.1186/s12976-017-0074-5
- [9] D. C. Montgomery, *Introduction to Statistical Quality Control*, 8th ed. Hoboken, NJ: John Wiley & Sons, 2019.
- [10] J. W. Buehler et al., "Framework for evaluating public health surveillance systems for early detection of outbreaks", *MMWR Morb. Mortal. Wkly. Rep.*, vol. 53, no. RR05, pp. 1-11, 2004.
- [11] Pan American Health Organization, "Technical guidelines for epidemiological surveillance of acute respiratory infections," PAHO/WHO, Washington, D.C., 2020.
- [12] E. S. Page, "Continuous inspection schemes", *Biometrika*, vol. 41, no. 1-2, pp. 100-115, 1954. doi: 10.1093/biomet/41.1-2.100
- [13] S. W. Roberts, "Control chart tests based on geometric moving averages", *Technometrics*, vol. 1, no. 3, pp. 239-250, 1959. doi: 10.1080/00401706.1959.10489860
- [14] T. Fawcett, "An introduction to ROC analysis", *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861-874, 2006. doi: 10.1016/j.patrec.2005.10.010
- [15] A. S. Buckridge et al., "Algorithms for rapid outbreak detection: a research synthesis", *J. Biomed. Inform.*, vol. 38, no. 2, pp. 99-113, 2005. doi: 10.1016/j.jbi.2004.11.007
- [16] Instituto Nacional de Estadística Honduras, "Encuesta Nacional de Demografía y Salud (ENDESA/MICS 2019)", INE, Tegucigalpa, 2020.

- [17]S. Morse, "Public health surveillance and infectious disease detection", *Biosecur Bioterror.*, vol. 10, no. 1, pp. 6-16, 2012. doi: 10.1089/bsp.2011.0088
- [18]D. M. Sosin, "Draft framework for evaluating syndromic surveillance systems", *J. Urban Health*, vol. 80, no. 2, Suppl. 1, pp. i8-i13, 2003, doi: 10.1007/PL00022309
- [19]K. M. Sullivan et al., "Monitoring drinking water quality in developing countries", *J Water Health*, vol. 5, no. 1, pp. 31-44, 2007. doi: 10.2166/wh.2006.043
- [20]World Health Organization, ICD-10: International Statistical Classification of Diseases and Related Health Problems, 10th revision. Geneva: WHO, 2016.
- [21]W. K. de Oliveira et al., "Establishing a national real-time surveillance system for respiratory diseases in Brazil", *Emerg. Infect. Dis.*, vol. 25, no. 4, pp. 675-682, 2019. doi: 10.3201/eid2504.171900
- [22]H. Gomez-Dantes et al., "Syndromic surveillance system for influenza in Mexico", *Salud Publica Mex.*, vol. 54, no. 3, pp. 307-315, 2012.
- [23]H. Burkom et al., "Role of data aggregation in biosurveillance detection strategies with applications from ESSENCE", *MMWR Morb. Mortal. Wkly. Rep.*, vol. 53, Suppl, pp. 67-73, 2004.
- [24]J. Shiffman, K. G. Tobey, and Y. Chen, "Advances in text mining for biosurveillance", *Comput. Math. Methods Med.*, vol. 2019, pp. 1-12, 2019. doi: 10.1155/2019/6712010
- [25]R. T. Milinovich et al., "Internet-based surveillance systems for monitoring emerging infectious diseases", *Lancet Infect. Dis.*, vol. 14, no. 2, pp. 160-168, 2014. doi: 10.1016/S1473-3099(13)70244-5
- [26]M. Paterson et al., "Pandemic influenza: a qualitative review of the literature", *Can J Infect Dis Med Microbiol*, vol. 22, no. 1, pp. 29-32, 2011. doi: 10.1155/2011/716921
- [27]D. A. Rasko et al., "The promise and potential of biosurveillance", *Annu Rev Biomed Data Sci*, vol. 1, pp. 175-195, 2018. doi: 10.1146/annurev-biodatasci-080917-013339
- [28]R. L. Brownstein et al., "Digital disease detection harnessing the web for public health surveillance", *N. Engl. J. Med.*, vol. 360, no. 21, pp. 2153-2157, 2009. doi: 10.1056/NEJMp0900702