

**AGRUPAMIENTO DE FENOTIPOS ESTABLECIDOS A PARTIR DE APRENDIZAJE  
DE MÁQUINA NO SUPERVISADO PARA PACIENTES CON SEPSIS NEONATAL NO  
CONFIRMADA Y MENOR A 33 SEMANAS SEGÚN TASA DE USO DE ANTIBIÓTICOS**

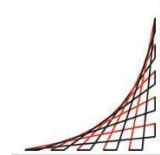
**Juan Carlos Beltrán Gasca**

**Tutor**

**Alvaro David Orjuela Cañón**



**Universidad del  
Rosario**



**ESCUELA  
COLOMBIANA  
DE INGENIERÍA  
JULIO GARAVITO**

**UNIVERSIDAD DEL ROSARIO  
ESCUELA COLOMBIANA DE INGENIERÍA JULIO GARAVITO  
PROGRAMA DE INGENIERÍA BIOMÉDICA  
BOGOTÁ D.C  
2025**

**AGRUPAMIENTO DE FENOTIPOS ESTABLECIDOS A PARTIR DE APRENDIZAJE  
DE MÁQUINA NO SUPERVISADO PARA PACIENTES CON SEPSIS NEONATAL NO  
CONFIRMADA Y MENOR A 33 SEMANAS SEGÚN TASA DE USO DE ANTIBIÓTICOS**

**Juan Carlos Beltrán Gasca**

Trabajo final de maestría presentado como requisito para optar al título de:

**Magister en Ingeniería Biomédica**

**UNIVERSIDAD DEL ROSARIO**

**ESCUELA COLOMBIANA DE INGENIERÍA JULIO GARAVITO**

**PROGRAMA DE INGENIERÍA BIOMÉDICA**

**BOGOTÁ D.C**

**2025**

## AGRADECIMIENTOS

Quiero expresar mi más sincero agradecimiento a la **Universidad del Rosario** y a la **Escuela Colombiana de Ingeniería Julio Garavito** por haber organizado y desarrollado este programa académico, el cual me brindó, como médico, la invaluable oportunidad de adquirir conocimientos fundamentales en aprendizaje de máquina y su aplicación en el ámbito clínico.

Gracias a esta formación, he podido comprender con mayor profundidad cómo las herramientas de inteligencia artificial pueden ser utilizadas para abordar problemas médicos complejos y relevantes, en especial aquellos en los que las decisiones clínicas requieren ser optimizadas con base en evidencia generada a partir de datos. Esta experiencia no solo ha ampliado mi perspectiva profesional, sino que también me ha motivado a seguir explorando la intersección entre la medicina y la ingeniería, con el propósito de contribuir a un cuidado más preciso, eficiente y humano.

Reitero mi gratitud por el compromiso académico e interdisciplinario que hizo posible este proceso formativo.

## RESUMEN

La sepsis neonatal causa mucha mortalidad y morbilidad sobre todo en la población de prematuros, aclarando que no hay consenso en su definición y que la sospecha se basa fundamentalmente en el cambio de la evolución del neonato examinado, lo que condiciona al médico tratante al inicio de antibióticos de manera empírica con amplia cobertura y con una duración muy variada. Estos dos últimos factores conllevan a mayor morbi- mortalidad. Insistir en esta asociación es muy importante por que como se demuestra en estudios recientes, aunque se ha logrado con el tiempo una disminución en los días de tratamiento, aún no se ha logrado disminuir el número de neonatos expuestos a los antibióticos empíricos. Se pretende utilizar una estrategia de aprendizaje de máquina para la búsqueda de esta asociación entre duración de antibióticos empíricos y muerte en una base de datos anónimos de libre acceso en la que todos los pacientes fueron evaluados y tratados por sospecha de sepsis temprana o sepsis tardía.

***Palabras claves:*** *Sepsis Neonatal, Tasa de Uso de Antibióticos (AUR), Aprendizaje de máquina, Mortalidad.*

## **ABSTRACT**

Neonatal sepsis causes significant mortality and morbidity, especially in the premature population. It should be noted that there is no consensus on its definition and that suspicion is primarily based on changes in the neonate's progress, which leads the treating physician to initiate empirical antibiotics with broad coverage and widely varying durations. These last two factors lead to increased morbidity and mortality. Emphasizing this association is very important because, as recent studies have shown, although a decrease in treatment days has been achieved over time, the number of neonates exposed to empirical antibiotics has not yet been achieved. We intend to use a machine learning strategy to search for this association between the duration of empirical antibiotics and death in an open-access, anonymous database in which all patients were evaluated and treated for suspected early or late sepsis.

**Keywords:** Neonatal Sepsis, Antibiotic Use Rate (AUR), Machine Learning, Mortality.

## TABLA DE CONTENIDO

<b>1.</b>	<b>INTRODUCCIÓN</b>	10
<b>2.</b>	<b>OBJETIVOS</b>	15
2.1	General	15
2.2	Específicos	15
<b>3.</b>	<b>MARCO TEORICO</b>	16
<b>4.</b>	<b>METODOLOGÍA</b>	21
4.1	Etapa 1: Establecimiento de base de datos	21
4.2	Etapa 2: Preprocesamiento de los Datos.	23
4.3	Etapa 3: Creación de los modelos de Aprendizaje No Supervisado y Validación de los resultados obtenidos.	24
4.4	Etapa 4: Métricas de calidad de agrupamiento	28
<b>5.</b>	<b>RESULTADOS</b>	29
<b>6.</b>	<b>DISCUSIÓN</b>	39
<b>7.</b>	<b>RECOMENDACIONES Y TRABAJOS FUTUROS</b>	40
<b>8.</b>	<b>CONCLUSIÓN</b>	41
<b>9.</b>	<b>REFERENCIAS</b>	42

## LISTA DE TABLAS

<i>Tabla I CLASIFICACIÓN DE EPISODIOS EN BASE DE DATOS ORIGINAL.</i>	22
<i>Tabla II VARIABLES</i>	25
<i>Tabla III RESUMEN COMPARATIVO</i>	28
<i>Tabla IV DESCRIPCIÓN DE LAS VARIABLES CATEGORICAS UTILIZADAS PARA EL AGRUPAMIENTO</i>	31
<i>Tabla V FALLECIDOS VS NO FALLECIDOS</i>	32
<i>Tabla VI CLUSTERS SEGÚN INDICES Y ALGORITMOS</i>	35
<i>Tabla VII CARACTERÍSTICAS DE LOS CLUSTERS</i>	37

## LISTA DE ILUSTRACIONES

<i>Ilustración 1. Variables numéricas</i>	33
<i>Ilustración 2. Clusters según K Means</i>	34
<i>Ilustración 3. Clusters según DBSCAN</i>	34
<i>Ilustración 4. Clusters según Jerárquico</i>	35
<i>Ilustración 5. Clusters según SOMNET</i>	35

## 1. INTRODUCCIÓN

La sepsis es una causa importante de morbilidad y de mortalidad a nivel global. La Organización Mundial de la Salud generó una resolución en 2017, en la cual se establece para los 194 miembros de las Naciones Unidas, la prioridad en la identificación y en el tratamiento de la misma. A partir de enero 21 del 2024 los nuevos criterios definen la sepsis en pediatría como la disfunción orgánica múltiple de los sistemas respiratorio, cardiovascular, de coagulación y/o neurológico en el contexto de infección probada o sospechada [1].

Es muy importante resaltar que la sepsis representa la principal causa de mortalidad en recién nacidos, siendo responsable de más de un millón de muertes neonatales en todo el mundo cada año, y los antibióticos son los medicamentos recetados con mayor frecuencia en las unidades de cuidado intensivo neonatal [2]. Pero a diferencia de lo establecido para sepsis en pediatría, a nivel neonatal (primeros 30 días de vida), la definición de sepsis neonatal aún requiere de estandarización para diagnósticos precisos, estudios epidemiológicos y síntesis de datos. Aclarando que cualquier definición de sepsis que incluya disfunción orgánica para determinar la sepsis neonatal, proporcionará imprecisiones que podrían afectar el verdadero impacto de una intervención cuando se aplica de manera apropiada. Por el contrario, una vez que se establece la estabilidad fisiológica (o al menos, un patrón constante de disfunción) en un neonato a término o prematuro, un cambio con respecto al valor inicial podría utilizarse de manera eficaz para inscribir a pacientes en riesgo en ensayos de intervención específicos [3].

En lo que si hay acuerdo es en que la sepsis neonatal tiene signos inespecíficos y conlleva consecuencias graves; como resultado, se inicia rápidamente un tratamiento antimicrobiano empírico en los recién nacidos sintomáticos con sospecha de sepsis después de obtener material biológico para cultivo [4]. Por lo que más del 95 % de los bebés ingresados en la Unidad de Cuidado Intensivo Neonatal reciben antibióticos empíricos en los primeros días postnatales, y solo entre el 1 y el 5 % tienen hemocultivos iniciales positivos [5]. El otro concepto sobre el cual hay en general acuerdo es el de la clasificación de la infección en relación al momento de aparición durante el periodo neonatal. La sepsis neonatal que se conoce como la sepsis de inicio temprano, suele aparecer antes de las 72 horas de vida, y la sepsis de inicio tardío, es la que aparece después de las 72 horas de vida. La distinción entre ambos tipos de infección se basa en los diferentes gérmenes, el modo de transmisión de la madre al recién nacido y el momento de aparición de los síntomas, y esta clasificación influye en el tratamiento y la terapia antibiótica. En la sepsis de inicio temprano, hay un paso transplacentario de bacterias de la madre al feto (transmisión vertical de la infección, como en las infecciones por *Listeria*) o una infección ascendente a través del tracto genital materno (como en las infecciones por *Streptococcus* del grupo B y *Escherichia coli*). La sepsis de inicio tardío es causada principalmente por organismos adquiridos horizontalmente del medio ambiente, ya sea de la comunidad o del ambiente hospitalario (en forma de infecciones estafilocócicas adquiridas de superficies, manos de trabajadores de la salud, catéteres venosos centrales u otros dispositivos), y también incluyen casos de neonatos que se presentan al departamento de emergencias con sospecha de sepsis [6].

El uso empírico de antibióticos de amplio espectro se asocia a diferentes efectos adversos: alteración de la colonización intestinal, aparición de cepas resistentes y aumento del riesgo de colonización por *Candida* y posterior candidiasis invasiva. Todos los antibióticos pueden alterar la colonización intestinal del paciente, promoviendo tanto la resistencia a los

antibióticos entre los organismos comensales normales como la aparición de otros patógenos [7]. El otro punto a considerar con la antibioticoterapia empírica es la duración de la misma ya que sigue habiendo dudas sobre la duración óptima de la terapia con antibióticos para la sepsis clínica, no demostrada microbiológicamente. Estudios de cohorte muestran una asociación entre la duración de la terapia con antibióticos empíricos y la mortalidad, la enterocolitis necrotizante y la duración de la estancia hospitalaria Cotten et al. [8] En dicho estudio se realizó un análisis de cohorte retrospectivo de 5693 lactantes con peso al nacer extremadamente bajo admitidos en 19 centros terciarios. De 5693 lactantes, 4039 sobrevivieron a más de 5 días, recibieron tratamiento antibiótico empírico inicial y tuvieron un hemocultivo inicial estéril a las 72 horas de vida. En un análisis multivariado ajustado por factores de riesgo, la duración prolongada de la terapia se asoció con mayores probabilidades de enterocolitis necrotizante o muerte o muerte sola. Cada día adicional de terapia con antibióticos se asoció con un aumento del 4 % en las probabilidades de enterocolitis necrotizante o muerte [9].

El conocimiento de todas estas asociaciones llevó a que se implementarán estrategias para el control en el uso de antibióticos y así es como en 2019, la Organización Mundial de la Salud (OMS) declaró que la Resistencia Antimicrobiana (AMR del inglés AntiMicrobial Resistance) es una de las 10 principales amenazas para la salud pública mundial que enfrenta la humanidad [10]. Anteriormente, en 2015, los Estados miembros de la OMS aprobaron un Plan de acción mundial sobre la AMR, que comprometía a los países a desarrollar planes de acción y acciones nacionales para abordar la AMR y reducir el uso inadecuado de antimicrobianos [11].

A nivel neonatal por ejemplo se planteó un estudio para evaluar las tendencias en el uso de antibióticos en una gran cohorte de bebés con peso extremadamente bajo al nacer (<1000 g; ELBW por sus siglas en inglés) ingresados en unidades de cuidado intensivo neonatal (UCIN) académicas y comunitarias en los EE. UU. durante un período de 13 años 2009-2021. La conclusión de los autores es: “Observamos una reducción sustancial en la duración del tratamiento con antibióticos a pesar de que no se produjo ningún cambio sustancial en la proporción de neonatos expuestos a antibióticos. Esto sugiere el éxito de las iniciativas de control orientadas a la duración del tratamiento con antibióticos y destaca la necesidad de mejorar los enfoques para identificar a los neonatos con bajo peso al nacer que corren mayor riesgo de infección” [12].

En los últimos años el uso de la inteligencia artificial y del aprendizaje de máquina ha aumentado de manera considerable en el campo de la medicina y especialmente en el área del cuidado crítico. Si a la inteligencia artificial la consideramos como un cerebro, entonces, el aprendizaje de máquina pasa a ser como el camino en el que el proceso mental es completado mientras se adquieren nuevas habilidades cognitivas.

En una revisión relativamente reciente sobre el uso del aprendizaje de máquina en sepsis neonatal se describe que los estudios a la fecha se enfocan en el uso del aprendizaje automático para la predicción de muerte y para diagnosticar posibles casos de sepsis. Los modelos de aprendizaje automático con mejor rendimiento fueron el bosque aleatorio y las redes neuronales [13]. Al mismo tiempo se insiste en la importancia de realizar estudios en donde se use el aprendizaje automático en sospecha de sepsis neonatal con hemocultivos negativos y el impacto según el uso de antibióticos. En este sentido algunos estudios han explorado el uso del análisis de conglomerados para evaluar el uso de antibióticos en neonatos

gravemente enfermos en un entorno de bajos recursos como el reportado por Benoni y colaboradores [14] en una cohorte retrospectiva que incluyó a todos los recién nacidos dados de alta con diagnóstico de probable sepsis neonatal de la UCIN del Hospital Católico St. Luke entre abril y julio de 2021. El protocolo de manejo antibiótico incluyó dos líneas según las directrices de la OMS y una tercera línea según las directrices internas del hospital. En el análisis de conglomerados, se estimó la distancia de Gower en función de los antibióticos empleados en las diferentes líneas y la duración de cada una. La mortalidad y la dificultad respiratoria fueron las variables de respuesta.

Se identificaron cuatro grupos de manejo de antibióticos. El grupo 1 (n=145, 74,4%) no tuvo cambios de antibióticos, utilizando solo la primera línea. El grupo 2 (n=26, 13,3%) tuvo un cambio de la primera a la segunda línea. El grupo 4 (n=9, 4,6%) tuvo dos cambios: de primera a segunda línea y luego a tercera. En el grupo 3 (n=15, 7,7%), los recién nacidos fueron cambiados de ceftriaxona/cloxacilina como segunda línea a antibióticos fuera de protocolo. No hubo diferencias en sexo, edad, peso al ingreso ni mortalidad bruta entre los grupos. El grupo 3 incluyó una mayor frecuencia de lactantes que no respiraron al nacer (53,3%, p=0,011) y que requirieron ventilación con bolsa (46,7%, p=0,039) en comparación con los otros grupos.

Dados los antecedentes y el creciente uso de modelos de aprendizaje de máquina, el presente trabajo busca hacer un análisis de datos disponibles asociados a sepsis neonatal para encontrar información relevante en el proceso de prescripción de medicamentos en el ejercicio médico de tratamiento con antibióticos ante la sospecha inicial de sepsis pero con hemocultivos negativos.

## **2. PROBLEMA DE INVESTIGACION**

### **2.1.1. Formulación de la pregunta**

**¿Qué fenotipos clínicos pueden ser identificados mediante algoritmos de aprendizaje de máquina no supervisado en una base de datos externa de neonatos menores de 33 semanas de edad gestacional, que recibieron antibioticoterapia ante sospecha de infección, con cultivos negativos para bacterias, hongos y virus, y sin criterios clínicos que justificaran la duración del tratamiento antibiótico?**

### 3. Descripción del problema

En las unidades de cuidado intensivo neonatal (UCIN), el tratamiento empírico con antibióticos es una práctica casi universal ante la sospecha clínica de sepsis por deterioro clínico. Esta estrategia, motivada por la alta letalidad de la condición, se aplica con especial frecuencia en neonatos prematuros, quienes presentan signos clínicos inespecíficos que dificultan el diagnóstico precoz. Sin embargo, la mayoría de estos pacientes tienen hemocultivos negativos, lo que sugiere que el diagnóstico de sepsis no se confirma en la gran mayoría de los casos.

A pesar de la ausencia de evidencia microbiológica de infección, muchos de estos neonatos continúan recibiendo tratamiento antibiótico por períodos prolongados. Esta práctica ha demostrado estar asociada a complicaciones importantes, incluyendo un mayor riesgo de enterocolitis necrotizante, disbiosis intestinal, candidiasis invasiva, aumento de la estancia hospitalaria y mayor mortalidad. En neonatos con peso extremadamente bajo al nacer, cada día adicional de tratamiento antibiótico se ha vinculado con un incremento significativo en la probabilidad de muerte o complicaciones graves, como lo evidencian estudios de cohorte multicéntricos.

Este uso prolongado e innecesario de antibióticos no solo pone en riesgo la salud inmediata del neonato, sino que también contribuye al desarrollo de resistencia antimicrobiana, considerada por la Organización Mundial de la Salud como una de las principales amenazas a la salud pública global. A pesar de este panorama crítico, actualmente no existen herramientas clínicas suficientemente precisas que permitan distinguir entre neonatos que realmente requieren tratamiento antibiótico y aquellos en quienes podría suspenderse de forma segura [17].

Dado el creciente uso de la inteligencia artificial en medicina, se presenta una oportunidad valiosa para aplicar modelos de aprendizaje automático a bases de datos clínicas neonatales con el fin de identificar patrones asociados al riesgo de mortalidad vinculado al uso prolongado de antibióticos, especialmente en casos donde los hemocultivos son negativos y el cuadro clínico no presenta la severidad suficiente que justifique la continuidad del tratamiento antimicrobiano.

#### 4. Justificación

La investigación sobre la sepsis neonatal es prioritaria debido a su impacto devastador en la salud global infantil. La sepsis constituye la principal causa de mortalidad en recién nacidos, responsable de más de un millón de muertes cada año. A pesar de su importancia, aún persisten múltiples vacíos en la comprensión, definición y tratamiento efectivo de esta condición, especialmente cuando se presentan hemocultivos negativos, lo cual ocurre en la mayoría de los casos. La falta de una definición estandarizada de sepsis neonatal y la heterogeneidad de sus signos clínicos dificultan la precisión diagnóstica, generando un uso extensivo de antibióticos de amplio espectro como tratamiento empírico [3].

Este enfoque, aunque comprensible, tiene consecuencias graves: altera la microbiota intestinal del neonato, favorece la aparición de patógenos resistentes y se asocia a complicaciones como la enterocolitis necrotizante y una mayor mortalidad [18]. La prolongación innecesaria de la terapia antimicrobiana, en ausencia de confirmación microbiológica, expone a los pacientes a riesgos evitables. Además, contribuye al avance de la resistencia antimicrobiana, identificada por la OMS como una de las principales amenazas para la salud pública mundial.

Frente a este panorama, el uso de inteligencia artificial y modelos de aprendizaje automático ofrece una oportunidad innovadora y poderosa para mejorar el proceso de toma de decisiones clínicas. Si bien ya existen estudios que exploran el uso de estos modelos para predecir mortalidad o identificar casos confirmados de sepsis, se requiere de su aplicación en escenarios de sospecha clínica con hemocultivos negativos y antibioticoterapia, una situación clínica frecuente que amerita decisiones rápidas pero fundamentadas [16]. Esta investigación se propone cubrir ese vacío, analizando datos clínicos relevantes para apoyar un uso más racional y personalizado de los antibióticos.

El impacto de esta investigación es amplio y significativo. Por un lado, permitirá optimizar el tratamiento antibiótico empírico, minimizando los efectos adversos y mejorando los desenlaces clínicos en neonatos. Por otro, contribuirá a las estrategias globales de control de la resistencia antimicrobiana, promoviendo un uso más prudente de estos fármacos. Asimismo, esta propuesta tiene un valor metodológico, al fomentar el uso de herramientas de inteligencia artificial en contextos clínicos complejos, como el de la sepsis neonatal.

Los principales beneficiarios de esta investigación son los recién nacidos, especialmente aquellos internados en unidades de cuidado intensivo, quienes podrían recibir tratamientos más precisos y seguros. También se beneficiarán los profesionales de la salud, al contar con herramientas de apoyo para la toma de decisiones clínicas más informadas. Finalmente, los sistemas de salud y la sociedad en su conjunto verán un beneficio al reducir la carga de la resistencia antimicrobiana y mejorar los estándares de cuidado neonatal.

## 5. OBJETIVOS

### *5.1 General*

Analizar datos de pacientes de neonatología a través del uso de técnicas de agrupamiento con modelos de aprendizaje automático de tipo no supervisado e información de los registros adquiridos en el Hospital de Niños de Filadelfia que se encuentran disponibles para su estudio.

### *5.2 Específicos*

1. Establecer una base de datos de trabajo a partir de información de pacientes de neonatología con sepsis y tratamiento con antibióticos y limpieza de la misma a través de técnicas estadísticas.
2. Determinar grupos de interés en los datos analizados a partir de algoritmos de agrupamiento tomados del aprendizaje automático de tipo no supervisado.
3. Validar el agrupamiento desde el punto de vista clínico a partir de expertos en el área de sepsis en pacientes de neonatología.

## 6. MARCO TEORICO

### 6.1. Sepsis neonatal: definición e importancia clínica

La sepsis neonatal es una de las principales causas de morbilidad y mortalidad en recién nacidos, especialmente en prematuros, siendo responsable de más de un millón de muertes neonatales cada año a nivel mundial [1][2]. A pesar de su impacto, no existe una definición estandarizada universalmente aceptada, lo que dificulta su diagnóstico preciso, la comparación entre estudios y la evaluación de intervenciones clínicas [3].

En neonatos, la sepsis se clasifica según su momento de aparición en sepsis de inicio temprano (antes de las 72 horas de vida) y sepsis de inicio tardío (después de las 72 horas), dado que los agentes causales, rutas de transmisión y estrategias terapéuticas varían entre ambas [6]. El diagnóstico se basa frecuentemente en signos clínicos inespecíficos, lo que conlleva al inicio de terapia antimicrobiana empírica inmediata en más del 95 % de los recién nacidos sintomáticos, a pesar de que solo entre el 1 y el 5 % presentan hemocultivos positivos [5].

### 6.2. Uso de antibióticos y consecuencias clínicas

El tratamiento empírico con antibióticos de amplio espectro es una práctica extendida en unidades de cuidado intensivo neonatal (UCIN) [37]. No obstante, su uso prolongado, especialmente en neonatos prematuros sin confirmación microbiológica de sepsis, ha sido asociado a efectos adversos relevantes: disbiosis intestinal, candidiasis invasiva, aparición de cepas resistentes y, de manera crítica, un aumento en el riesgo de enterocolitis necrotizante, prolongación de la estancia hospitalaria y mortalidad [7][8][9].

Estudios retrospectivos han mostrado que en neonatos con peso extremadamente bajo al nacer (<1000 g), cada día adicional de tratamiento antibiótico en ausencia de infección confirmada aumenta el riesgo de complicaciones graves en un 4 % a 7.3 %, incluyendo deterioro neurológico y muerte [16][17][18]. Esta evidencia ha motivado iniciativas globales orientadas al control del uso de antimicrobianos, como el Plan de Acción Mundial contra la Resistencia Antimicrobiana de la OMS [10][11].

### 6.3. Tasa de Uso de Antibióticos (AUR)

Para cuantificar la exposición relativa a antibióticos, se utiliza la **Tasa de Uso de Antibióticos (AUR, por sus siglas en inglés)**, definida como el número de días de exposición a antimicrobianos dividido por la duración total de la hospitalización. Esta métrica ha demostrado ser un indicador útil del riesgo acumulativo asociado al uso prolongado de antibióticos en neonatos sin evidencia microbiológica de sepsis [16][17].

$$AUR = \frac{\text{días de antibioticos}}{\text{días de estancia}} \times 100 \quad (1)$$

### 6.4. Inteligencia artificial y aprendizaje de máquina en medicina neonatal

La inteligencia artificial (IA), y en particular el aprendizaje de máquina (machine learning), ha adquirido relevancia en medicina crítica por su capacidad de analizar grandes volúmenes de datos clínicos y detectar patrones complejos. En el contexto de la sepsis neonatal, los modelos más

prometedores han sido los bosques aleatorios y las redes neuronales artificiales, utilizados para predecir mortalidad o identificar casos de sepsis [13].

Sin embargo, existe un vacío importante: no se han desarrollado modelos dirigidos específicamente a neonatos con hemocultivos negativos, donde la decisión de continuar o suspender antibióticos es más incierta. En estos casos, los algoritmos de aprendizaje no supervisado —como K-Means, DBSCAN, clustering jerárquico y SOM (Self-Organizing Maps)— permiten identificar subgrupos (fenotipos clínicos) sin requerir etiquetas diagnósticas previas, siendo herramientas útiles para la estratificación del riesgo clínico y la optimización terapéutica [20][21][22][23].

**K-Means** es un algoritmo de agrupamiento ampliamente utilizado que organiza los datos en  $k$  clústeres o grupos, de modo que los elementos dentro de un mismo clúster sean lo más similares posible entre sí, y al mismo tiempo, lo más diferentes posible de los elementos pertenecientes a otros grupos. El algoritmo comienza seleccionando aleatoriamente  $k$  puntos como centros iniciales, conocidos como centroides. A partir de allí, cada punto del conjunto de datos se asigna al centroide más cercano, utilizando generalmente la distancia euclidiana como métrica de similitud. Una vez realizada esta asignación, se actualiza la posición de cada centroide calculando el promedio de todos los puntos que le han sido asignados. Este proceso de asignación y actualización se repite de manera iterativa hasta que los centroides dejan de cambiar significativamente o se alcanza un número máximo de iteraciones.

Los elementos fundamentales del algoritmo incluyen el número  $k$  de clústeres, que debe ser definido previamente; los centroides, que representan el centro geométrico de cada grupo; la distancia euclidiana, empleada como criterio de cercanía entre puntos; y la inercia, que mide la compacidad de los clústeres como la suma de las distancias cuadradas entre los puntos de cada grupo y su centroide. K-Means tiene como objetivo minimizar esta inercia, optimizando así la coherencia interna de cada agrupación [20].

Entre sus principales ventajas se encuentra su sencillez conceptual, facilidad de implementación y eficiencia computacional, lo que lo hace especialmente útil en conjuntos de datos grandes. Además, su naturaleza intuitiva lo ha convertido en una herramienta ampliamente adoptada en diversas áreas de análisis de datos. No obstante, presenta limitaciones importantes: requiere conocer de antemano el número óptimo de clústeres, es sensible a la presencia de valores atípicos y al escalado de las variables, puede converger a mínimos locales debido a su inicialización aleatoria y tiende a identificar clústeres de forma esférica y tamaño similar, lo que puede no ajustarse a la verdadera estructura de los datos.

**DBSCAN** (Density-Based Spatial Clustering of Applications with Noise) es un algoritmo de agrupamiento basado en densidad, diseñado para identificar grupos de puntos que se encuentran densamente conectados entre sí, al tiempo que detecta y separa aquellos que no pertenecen a ningún grupo, considerados como ruido o valores atípicos (*outliers*). A diferencia de métodos como K-Means, DBSCAN no requiere especificar de antemano el número de clústeres, y es capaz de detectar estructuras de agrupamiento con formas arbitrarias, lo que lo hace especialmente útil en escenarios donde los datos no se distribuyen de manera esférica. Además, muestra una mayor robustez frente al ruido y a los puntos aislados.

El funcionamiento de DBSCAN se basa en dos parámetros fundamentales. El primero,  $\epsilon$  (*epsilon*), define el radio de vecindad dentro del cual se evalúa la densidad local de puntos; el segundo, *min\_samples*, establece el número mínimo de puntos que deben encontrarse dentro de esa vecindad para que un punto sea considerado un "núcleo". A partir de estos criterios, el algoritmo clasifica cada observación como un punto núcleo (si cumple con el umbral de densidad), un punto frontera

(si se encuentra dentro del radio  $\epsilon$  de un núcleo pero no alcanza el número mínimo de vecinos) o un punto ruido (si no cumple ninguno de los criterios anteriores y no pertenece a ningún clúster) [21].

Aunque DBSCAN presenta ventajas importantes, como su capacidad para identificar clústeres complejos y su tolerancia al ruido, también enfrenta ciertas limitaciones. La selección adecuada de los parámetros  $\epsilon$  y *min\_samples* puede ser difícil y altamente dependiente de la estructura de los datos. Además, su rendimiento disminuye cuando los clústeres tienen densidades significativamente distintas, y su escalabilidad es limitada en conjuntos de datos muy grandes, aunque existen versiones optimizadas como HDBSCAN que abordan este problema.

**Los algoritmos jerárquicos** son un subconjunto de técnicas de agrupamiento (clustering) que crean una estructura de árbol o dendrograma para representar las relaciones entre los datos. Su objetivo es construir una jerarquía de grupos, donde cada nivel representa una resolución diferente del agrupamiento. Los Tipos principales de algoritmos jerárquicos son: Aglomerativos (Bottom-Up): Comienzan con cada elemento de los datos como un grupo individual. Los grupos se fusionan gradualmente en niveles superiores hasta formar un único clúster que engloba todos los datos. Divisivos (Top-Down): Comienzan con un único grupo que contiene todos los datos. Gradualmente, se dividen en subgrupos más pequeños hasta que cada elemento esté en su propio clúster. Este enfoque es menos común debido a su complejidad computacional.

La construcción de la jerarquía en estos algoritmos se basa en una métrica de similitud o distancia que define cómo se mide la relación entre los datos. Algunas de las métricas más comunes son: Distancia Euclidiana: Mide la separación directa entre dos puntos en un espacio multidimensional. Distancia Manhattan: Suma de las diferencias absolutas entre las coordenadas. Coeficiente de correlación: Evalúa la correlación entre los puntos. Para determinar cómo se combinan los clústeres, se utilizan métodos de enlace, entre los cuales destacan: Enlace Simple: Considera la distancia más corta entre dos elementos de diferentes clústeres. Enlace Completo: Usa la distancia más larga entre dos elementos de diferentes clústeres. Enlace Promedio: Calcula la distancia promedio entre todos los pares de elementos de dos clústeres. Centroide: Utiliza el centroide de un grupo como representante del clúster.

Un dendrograma es una representación gráfica de las relaciones jerárquicas entre los datos. Los nodos del árbol representan los datos individuales o grupos de datos. Las ramas muestran cómo se fusionan o dividen los clústeres en diferentes niveles de la jerarquía. Dentro de las ventajas se mencionan que no requiere especificar el número de clústeres de antemano, proporcionando una representación visual clara de las relaciones entre los datos (dendrograma) y es útil tanto para datos pequeños como medianos. Las desventajas incluyen la alta complejidad computacional, especialmente en conjuntos de datos grandes. La sensibilidad al ruido y a los datos atípicos, y que las decisiones tempranas en el proceso de agrupamiento (en el caso de métodos aglomerativos) pueden afectar significativamente los resultados finales [22].

**Los (SOM)**, también conocidos como Mapas de Kohonen (en honor a su creador Teuvo Kohonen), son un tipo de red neuronal que se utiliza en el aprendizaje no supervisado para la reducción de dimensionalidad y la agrupación (clustering) de datos. Su principal característica es que proyectan datos de alta dimensión en un espacio de menor dimensión, generalmente un espacio bidimensional, preservando las relaciones topológicas de los datos originales [23].

Un mapa consta de dos componentes principales: Capa de entrada: Representa los datos de entrada en forma de vectores de características. Cada vector de entrada tiene (n) dimensiones, donde (n) es el número de variables en el conjunto de datos. Capa de salida (o mapa): Consiste en una cuadrícula de neuronas organizadas en una estructura bidimensional (aunque también pueden tener más

dimensiones). Cada neurona tiene un vector de pesos asociado, que tiene la misma dimensión que los vectores de entrada.

El funcionamiento de un SOM se basa en los siguientes pasos: Inicialización: Los pesos de las neuronas del mapa se inicializan aleatoriamente o con valores pequeños cercanos al origen. Competencia: Para cada vector de entrada, se calcula la distancia (generalmente la distancia euclidiana) entre el vector de entrada y los pesos de todas las neuronas. La neurona cuya distancia sea menor se denomina Best Matching Unit (BMU), o Unidad de Mejor Coincidencia. Actualización: Los pesos de la BMU y de las neuronas vecinas se actualizan para aproximarse al vector de entrada, utilizando los pesos de la neurona, en el tiempo, tasa de aprendizaje, que decrece con el tiempo, función de vecindad, que decrece con la distancia entre la BMU y la neurona y el vector de entrada. Repetición: Este proceso se repite para todos los vectores de entrada durante un número determinado de iteraciones o hasta que el sistema converja.

Como propiedades de los SOM se mencionan la preservación de la topología de los datos de entrada, lo que significa que los datos similares en el espacio de entrada estarán cerca unos de otros en el mapa, la reducción de dimensionalidad facilitando su visualización y análisis, el clustering agrupando datos similares en regiones específicas del mapa, lo que permite identificar patrones subyacentes en los datos y el aprendizaje competitivo en el que las neuronas compiten entre sí para representarse a sí mismas como las más cercanas a los datos de entrada.

Las ventajas incluyen que el algoritmo es fácil de entender e implementar, genera representaciones gráficas útiles para analizar datos complejos, puede manejar datos de diferentes dominios y escalas y es ideal para problemas donde los datos no están etiquetados. Las Limitaciones como que no es adecuado para conjuntos de datos extremadamente grandes debido a su alta complejidad computacional, los resultados pueden depender del tamaño del mapa, la tasa de aprendizaje y la función de vecindad y el elegir incorrectamente las dimensiones del mapa puede llevar a una pérdida de información o interpretaciones incorrectas.

Para evaluar la calidad de los agrupamientos obtenidos, se emplearon como métricas de calidad del agrupamiento, tres índices ampliamente reconocidos en la literatura: **índice de la silueta**, **índice de Calinski-Harabasz** y **índice de Davies-Bouldin** [24].

El **índice de la silueta** permite valorar simultáneamente la cohesión intra-cluster y la separación inter-cluster [24]. Para cada observación  $i$ , se calcula como:

$$s(i) = \frac{b(i)-a(i)}{\max(a(i),b(i))} \quad (2)$$

donde  $a(i)$  corresponde a la distancia media entre  $i$  y los demás elementos de su propio cluster, y  $b(i)$  es la mínima distancia media entre  $i$  y los elementos de otros clusters. Los valores de  $s(i)$  oscilan entre -1 y 1, donde valores cercanos a 1 reflejan una adecuada asignación del punto a su cluster, valores próximos a 0 indican ambigüedad en la asignación, y valores negativos sugieren una posible mala asignación. Se utiliza habitualmente el promedio global de  $s(i)$  como medida resumen de la estructura del clustering.

El **índice de Calinski-Harabasz** (también denominado *variance ratio criterion*) cuantifica la relación entre la dispersión inter-cluster y la dispersión intra-cluster [24]. Su formula es la siguiente:

$$CH = \frac{Tr(B_k)}{k-1} \div \frac{Tr(W_k)}{n-k} \quad (3)$$

donde  $B_k$  y  $W_k$  representan, respectivamente, las matrices de dispersión entre clusters y dentro de los clusters,  $k$  es el número de clusters y  $n$  el número total de observaciones. Valores elevados del índice indican clusters bien diferenciados y compactos, y el valor óptimo del número de clusters se asocia con el máximo del índice.

El **índice de Davies-Bouldin** evalúa la calidad del clustering en función de la similitud entre los clusters. Se define como:

$$DB = \frac{1}{k} \sum_{i=1}^k \max_{j \neq i} \left( \frac{s_i + s_j}{d_{ij}} \right) \quad (4)$$

donde  $s_i$  es la dispersión interna del cluster  $i$  y  $d_{ij}$  corresponde a la distancia entre los centroides de los clusters  $i$  y  $j$ . Valores bajos de este índice reflejan una mejor separación entre clusters y mayor compacidad interna.

### **6.5. Aplicación del aprendizaje no supervisado en la sepsis neonatal**

El presente trabajo se enfoca en analizar una cohorte de neonatos menores de 33 semanas con sospecha de sepsis y hemocultivos negativos, utilizando algoritmos de agrupamiento no supervisado. Se busca determinar si existen patrones clínicos (fenotipos) que se asocien a una mayor exposición antibiótica (AUR) y, por tanto, a un mayor riesgo de mortalidad. La hipótesis es que un subgrupo de pacientes podría estar recibiendo tratamiento innecesario, y que su identificación permitiría una reducción segura del uso de antibióticos en esta población vulnerable

## 7. METODOLOGÍA

Para el desarrollo del presente proyecto se proponen tres etapas que se desarrollan de forma secuencial:

### *7.1 Etapa 1: Establecimiento de base de datos*

El conjunto de datos es una cohorte completamente anónima de 986 neonatos evaluados ante la sospecha de infección en el Hospital de Niños de Filadelfia [18]. Los datos se recopilaron entre septiembre de 2014 y febrero de 2018. Estos datos según es reportado en el artículo original se obtuvieron de la siguiente manera:

- **Identificación de casos:** Mediante lenguaje de consulta estructurado fueron identificadas las órdenes de hemocultivo de los neonatos ingresados en la unidad de cuidados intensivos neonatales (UCIN) seguidas de una orden de antibiótico en las 24 horas siguientes. Los investigadores revisaban estos datos para confirmar que correspondían a evaluaciones reales de sepsis, descartando errores o repeticiones.
- **Recopilación de datos:** Una vez confirmada una evaluación de sepsis, un script transfería datos básicos (resultados de cultivos, laboratorio, signos vitales, antibióticos y datos demográficos) para análisis preliminares.
- **Duración de Antibioticoterapia:** Tras un seguimiento adecuado, se definió para cada caso la duración del antibiótico, según los cultivos y la condición clínica al momento del estudio por sospecha de sepsis. como se indica en Tabla I.

Tabla I  
DURACION DE ANTIBIOTICOS EN BASE DE DATOS ORIGINAL.

<b>Criterio</b>	<b>Condición Clínica</b>	<b>Descripción breve</b>
1	Bacteriana confirmada	Cultivo positivo; $\geq 5$ días ( $\geq 120$ h) de antibiótico.
2	No infecciosa	Cultivo negativo; $\leq 72$ h de antibiótico.
3	Sepsis clínica	Cultivo negativo; $\geq 5$ días ( $\geq 120$ h) de antibiótico.
4	Infección viral	Cultivo positivo solo viral; negativo para bacterias.
5	Infección fúngica	Cultivo positivo solo fúngico; negativo para bacterias y virus.
6	Otros	Cultivo negativo; $> 72$ h pero $< 120$ h de antibiótico.

En este punto, 1,3,4 y 5 descritos en la Tabla I fueron eliminados ya que la continuidad de los antibióticos fue condicionada por la positividad de los cultivos en 1, por el importante compromiso clínico en 3 y para 4 y 5 la pronta suspensión de los antibióticos se dio al documentar sepsis por agente viral o por agente fúngico respectivamente y que por lo tanto no requieren de manejo antibiótico. El estudio entonces se centra en los neonatos incluidos en 2 (No infecciosa) y 6 (Otros) en donde los hemocultivos fueron reportados como negativos y la condición de los neonatos no cumplió con criterios de severidad como en 3 de sepsis clínica, pero se dio continuidad a la prescripción de los antibióticos hasta tres y cuatro días respectivamente. El resto de las variables incluidas están descritas en la Tabla II.

La variable duración de estancia en horas de la base de datos original junto con las horas de antibioticoterapia a pesar de hemocultivos negativos de acuerdo con los criterios 2 y 6 por paciente menor de 33 semanas de edad gestacional, se convirtieron en una nueva variable denominada Tasa de Uso de Antibióticos (Antibiotic Use Rate por su sigla en inglés AUR) de acuerdo con la Ecuación 1, definida como el número de días que un neonato está expuesto a uno o más agentes antimicrobianos, dividido entre la duración total de la hospitalización. Esto es importante, dado que diversos estudios han evidenciado una asociación entre el uso prolongado de antibióticos en recién nacidos y un mayor riesgo de resultados adversos, incluso en ausencia de infecciones confirmadas. Un estudio nacional que incluyó a 8824 lactantes sin sepsis de inicio temprano, sepsis de inicio tardío ni enterocolitis necrosante encontró que las tasas más altas de uso de antibióticos (AUR) se relacionaron significativamente con un mayor riesgo de mortalidad o morbilidad mayor, definida por alteraciones persistentes en neuroimagen, retinopatía del prematuro en estadio avanzado y enfermedad pulmonar crónica [14]. De forma similar, otro análisis que incluyó a 14207 lactantes sin evidencia de infección comprobada mostró que el 41 % recibió antibióticos durante 4 a 7 días en la primera semana de vida, y cada día adicional de antibióticos se asoció con un incremento del 7,3 % en el riesgo de muerte o morbilidad severa, particularmente en prematuros de muy bajo peso al

nacer con bajo riesgo de infección temprana [15]. Por último, un estudio de cohorte retrospectivo en 1373 lactantes menores de 29 semanas de gestación ingresados en UCIN reveló que aquellos expuestos a las tasas más altas de antibióticos presentaron una probabilidad significativamente mayor de muerte o deterioro importante del neurodesarrollo (ORa: 7,44; IC 95 %: 4,55–12,2) [16].

## ***7.2 Etapa 2: Preprocesamiento de los Datos.***

El aprendizaje no supervisado, como el agrupamiento, se basa en encontrar patrones ocultos en los datos sin etiquetas previas. Para garantizar resultados interpretables y útiles, el preprocesamiento previo es fundamental. A continuación, se describen las características del preprocesamiento aplicadas.

Durante la etapa de preparación de los datos, las variables del conjunto fueron clasificadas en dos grupos principales. Por un lado, se identificaron las variables categóricas, que incluyen aquellas de tipo binario o cualitativo, como el sexo o la presencia de una línea central. Por otro lado, se reconocieron las variables numéricas, correspondientes a datos continuos como la edad gestacional, el peso al nacer o los días sin necesidad de intubación. Esta separación permitió aplicar distintos tipos de transformación según la naturaleza de cada variable. En el caso de las variables categóricas, se convirtieron en columnas binarias para facilitar su uso en modelos matemáticos, omitiendo intencionalmente una de las categorías posibles para evitar problemas de redundancia y sobreajuste relacionados con la multicolinealidad.

Dentro de un proceso de análisis de datos con modelos de aprendizaje automático, es fundamental aplicar una transformación a las variables numéricas para evitar que las diferencias de escala influyan en los resultados. Una técnica común es la normalización, que ajusta los valores de cada variable para que se ubiquen dentro del intervalo (0, 1). Esto se logra calculando el valor mínimo y máximo de cada variable y reescalando los datos en función de estos extremos. Este procedimiento es especialmente útil en algoritmos sensibles a la magnitud de las variables, como los métodos de agrupamiento o los basados en distancia. Una forma habitual de implementar esta normalización es mediante una transformación conocida como MinMax scaling, ampliamente empleada en la práctica del aprendizaje automático [17].

Otra operación realizada durante el preprocesamiento consistió en la codificación de variables categóricas ( sexo, mortalidad general en 30 días, presencia de línea venosa central, línea arterial umbilical, temperatura y comorbilidad por *ivh*\_derivación ). Para evitar que los modelos interpreten un orden arbitrario entre las categorías, se optó por una codificación binaria conocida como *one-hot encoding*, que transforma cada categoría en una columna distinta con valores 0 o 1. Este enfoque evita la introducción de relaciones ficticias entre categorías y es ampliamente recomendado en el procesamiento de datos categóricos para modelos de aprendizaje automático. En este caso, se utilizó la implementación proporcionada por la biblioteca Scikit-Learn [19].

Después de aplicar la codificación a las variables categóricas, se realizó un resumen de la información transformada. Este paso resulta útil tanto para verificar que el preprocesamiento se haya llevado a cabo correctamente como para comprender la nueva estructura del conjunto de datos que será utilizado en el entrenamiento de los modelos no supervisados. Cabe destacar que, en el grupo de recién nacidos con menos de 33 semanas de gestación, no se identificaron datos faltantes en la base de datos original.

### 7.3. Etapa 3: Creación de los modelos de Aprendizaje No Supervisado y Validación de los resultados obtenidos.

Para identificar grupos representativos dentro del conjunto de datos, se emplean algoritmos de agrupamiento basados en medidas de similitud. Entre los métodos seleccionados para este análisis se incluyen técnicas clásicas como k-means, DBSCAN y algoritmos jerárquicos que utilizan dendrogramas, como el método linkage. Adicionalmente, se consideran enfoques más avanzados basados en redes neuronales artificiales, como los mapas autoorganizados (Self-Organizing Maps, SOM). Para el entrenamiento de cada uno de estos modelos, se definirán los hiper parámetros correspondientes, teniendo en cuenta tanto las métricas de similitud como los indicadores de calidad del agrupamiento, con el objetivo de identificar las configuraciones que mejor representen la estructura subyacente de los datos. Utilizamos los datos para el desarrollo del modelo y aplicamos tres métricas internas de agrupamiento para comparar la consistencia, separación y distribución de los conglomerados de los algoritmos. El coeficiente de Silueta mide la consistencia de los conglomerados, evaluando la similitud de una observación con su conglomerado en comparación con los demás; los valores más altos indican una mejor coincidencia con su propio conglomerado. El índice de Davies-Bouldin mide la separación de conglomerados promediando la similitud máxima entre todos los conglomerados; los valores más bajos indican conglomerados más distintos. El índice de Calinski-Harabasz es la razón entre la dispersión acumulada entre conglomerados y la dispersión dentro de los conglomerados; los valores más altos indican conglomerados mejor definidos.

Clasificamos los cuatro algoritmos utilizando cada una de las tres métricas; las clasificaciones más bajas indican un mejor rendimiento; la clasificación general fue la suma de las clasificaciones individuales. 44]. Además de la validación de agrupamientos basada en la distancia, calculamos la entropía, una medida de separación de clases, que puede ser informativa sobre la precisión de la diferenciación de los agrupamientos al medir la precisión del fenotipo estimado de cada paciente. La entropía varía de 0 a 1; valores cercanos a 1 indican una alta probabilidad de que los pacientes pertenezcan a una sola clase. Utilizamos la fórmula proporcionada por Asparouhov:

$$E = 1 + \frac{1}{N \log(k)} \left( \sum_{i=1}^N \sum_{k=1}^K P(C = k | U_i) \log(p(C = k | U_i)) \right) \quad (5)$$

donde C es la variable latente, K es el número de clases, N es el tamaño de la muestra,  $U_i$  es el vector de todas las variables indicadoras de clase latente y las probabilidades  $P(C = k | U_i)$  se calculan a partir del modelo estimado. Cuanto mayor sea la entropía, más clara será la identificación de la clase latente. El valor de la entropía está entre 0 y 1. Una entropía con valores cercanos a 1 indica una clara separación de las clases.

Tabla II  
VARIABLES

<b>Variables Categóricas</b>	<b>Definición</b>
Sexo	Sexo biológico, codificado como 0 femenino y 1 masculino
Mortalidad_general_en_30_días	Indica si el niño falleció dentro de los 30 días posteriores a la evaluación de sepsis por cualquier motivo (0 no, 1 sí).
Línea venosa central	Indica si había una vía venosa central (p. ej., una vía venosa umbilical o un catéter central de inserción periférica) en el momento de la evaluación de la sepsis (0 no, 1 sí).
Línea arterial umbilical	Indica si había un catéter arterial umbilical en el momento de la evaluación de la sepsis (0 no, 1 sí).
Temperatura	Indica si había distermia en el momento de la evaluación de la sepsis (0 no, 1 sí).
Comorbilidad_ivh_o_derivación	Presencia de hemorragia intraventricular o derivación ventrículo-peritoneal detectada en cualquier momento antes de la evaluación de sepsis (0 no, 1 sí).
<b>Variables Numéricas</b>	<b>Definición</b>
Edad gestacional al nacer (semanas)	Gestación posmenstrual al nacer, registrada en número completo de semanas.
Peso al nacer (kg)	Peso en kilogramos medido al momento del nacimiento
Edad_de_inicio_en_días	Edad del bebé en días de vida al inicio del estudio con hemocultivo.
Días_libres_de_intubación	Número de días que no estuvo intubado en los 28 días posteriores a la evaluación de sepsis. El estado de intubación en el último día de observación (p. ej., en el caso de los neonatos fallecidos) se mantuvo durante el resto del período de observación de 28 días.
Días_libre_de_inotrópicos	Número de días que el lactante no recibió soporte inotrópico en los 28 días posteriores a la evaluación de sepsis. El estado de soporte inotrópico en el último día de observación (p. ej., en el caso de los lactantes fallecidos) se mantuvo durante el resto del período de observación de 28 días.
AUR	Indica la Tasa de Uso de Antibióticos definida como el número de días que un neonato está expuesto a uno o más agentes antimicrobianos,

	dividido entre la duración total de la hospitalización.
--	---

### 7.3.1. Algoritmo de K-Means

En el modelo de agrupamiento, se aplicó el algoritmo K-Means con el objetivo de identificar cuatro grupos distintos dentro del conjunto de datos. Para garantizar la reproducibilidad de los resultados en futuras ejecuciones, se estableció una condición de aleatoriedad controlada. El análisis se llevó a cabo sobre un conjunto de datos previamente preprocesado, en el que las variables fueron normalizadas y las categorías codificadas en formato binario.

Una vez entrenado el modelo, se asignó a cada observación una etiqueta numérica que indica el grupo al que pertenece, utilizando valores enteros del 0 al 3, correspondientes a los cuatro clústeres definidos. Estas etiquetas permiten clasificar cada fila del conjunto de datos según su similitud con los patrones detectados.

Posteriormente, se identificaron las coordenadas de los centroides de cada clúster en el espacio de variables transformadas, lo que permite interpretar las características promedio de cada grupo. Además, se obtuvo una medida de la inercia total del modelo, que corresponde a la suma de las distancias cuadradas entre cada punto y el centroide de su clúster respectivo. Esta métrica cuantifica la compacidad de los grupos formados: valores menores indican una mayor cohesión interna, aunque un exceso de clústeres puede llevar al sobreajuste del modelo.

### 7.3.2. Algoritmo *Density-Based Spatial Clustering of Applications with Noise* (DBSCAN)

En el desarrollo de los experimentos, se aplicó el algoritmo DBSCAN utilizando una distancia máxima de 0,55 entre puntos para definir la vecindad, y un valor mínimo de 15 observaciones dentro de esa vecindad para que un punto sea considerado como núcleo. El análisis se realizó sobre un conjunto de datos previamente preprocesado, que incluía limpieza, normalización y ausencia de valores nulos. Al ejecutar el algoritmo sobre estos datos, se asignó una etiqueta de agrupamiento a cada observación. Los valores numéricos de estas etiquetas identifican los diferentes clústeres detectados, mientras que aquellas observaciones que no cumplieron con los criterios de densidad fueron clasificadas con un valor especial que indica que se consideran puntos ruido o atípicos.

### 7.3.3. Algoritmos jerárquicos – *Linkage*

En nuestra base de datos para calcular la matriz de similitud se utilizó la similaridad por coseno. Esta métrica obtiene similitudes en los vectores de entrada empleando la función coseno para su cálculo. La similitud del coseno mide el grado de similitud entre dos vectores en un espacio multidimensional, ignorando su magnitud. Su rango va de 0 a 1. Como resultado se obtiene una matriz de similitud cuadrada donde cada elemento ((i, j)) representa la similitud entre la fila (i) y la fila (j). Luego se convierte la matriz de similitud en un DataFrame de pandas para facilitar su visualización o análisis. Esto no afecta directamente el cálculo del algoritmo, pero sirve para inspeccionar los datos.

La construcción de la estructura jerárquica de agrupamiento se realizó a partir de una matriz de similitud previamente generada, la cual representa el grado de similitud entre las observaciones del conjunto de datos. Para este proceso, se utilizó un enfoque aglomerativo, empleando el método de Ward, que busca minimizar la varianza interna dentro de cada grupo en cada paso de fusión. Este método tiende a producir clústeres más homogéneos al evitar la formación de grupos con alta dispersión interna.

El resultado de esta operación es una matriz que describe paso a paso cómo se fueron fusionando los distintos grupos a lo largo del proceso. Cada fila de esta matriz representa una fusión e incluye información como los índices de los dos clústeres que se agruparon, la distancia entre ellos en el momento de la fusión y el número total de elementos que componen el nuevo grupo formado.

Para visualizar las relaciones jerárquicas obtenidas, se generó un dendrograma, es decir, un gráfico en forma de árbol que permite observar de manera clara la secuencia de fusiones y la distancia a la que ocurrieron. El gráfico fue personalizado para mejorar su legibilidad, incluyendo etiquetas en los ejes y un tamaño de figura adecuado para visualizar todas las observaciones.

Finalmente, para determinar a qué grupo pertenece cada observación, se estableció un umbral de distancia específico sobre el dendrograma. Al aplicar este criterio, se segmentó la estructura jerárquica en diferentes clústeres, asignando a cada elemento del conjunto de datos una etiqueta que indica el grupo al que fue clasificado. Estos resultados permiten analizar la distribución interna de los datos y la coherencia de los clústeres formados.

#### 7.3.4. Mapas Auto-Organizados - *Self-Organizing Maps* (SOM)

Se utilizó una herramienta especializada para implementar y entrenar un **Mapa Autoorganizado de Kohonen (Self-Organizing Map, SOM)**, diseñada específicamente para este tipo de análisis no supervisado. En este caso, se configuró un mapa bidimensional con una cuadrícula de 10 por 10 neuronas, lo que representa un total de 100 nodos, sobre los cuales se distribuyeron los datos del conjunto ya preprocesado. Cada neurona del mapa fue entrenada para aprender un vector de pesos que representa un patrón característico dentro de los datos originales, permitiendo así una organización topológica de las observaciones en un espacio reducido.

Durante el entrenamiento, la red ajustó los pesos de las neuronas de manera iterativa, reorganizando el espacio de representación de forma que las observaciones similares quedaran proyectadas en regiones próximas del mapa. Una vez finalizado el proceso de aprendizaje, se generaron representaciones gráficas que muestran cómo se distribuye cada variable del conjunto de datos sobre el mapa. Estas visualizaciones permiten observar la influencia relativa de cada variable en las distintas áreas del SOM, facilitando la interpretación de las relaciones entre las características.

Posteriormente, cada observación fue proyectada sobre su **neurona más representativa**, es decir, la que tiene el vector de pesos más cercano a dicha observación en el espacio de características. Se almacenó la posición de esta neurona ganadora y se incorporó una nueva columna al conjunto de datos indicando a qué nodo fue asignada cada muestra. A partir de esta información, fue posible calcular cuántas observaciones se agruparon en cada neurona, lo cual permite identificar las zonas del mapa con mayor densidad de datos y explorar los patrones dominantes en esas regiones. Además, se examinó el vector de pesos correspondiente a la primera neurona del mapa como ejemplo representativo del tipo de información aprendida durante el entrenamiento.

Como complemento al análisis con SOM, se entrenó un modelo de agrupamiento clásico mediante partición, configurado para identificar cuatro grupos distintos dentro del mismo conjunto de datos.

Tras el entrenamiento, se asignó a cada muestra una etiqueta de clúster correspondiente, lo que permitió comparar los resultados de ambos enfoques y evaluar la consistencia entre la organización autoorganizada y el agrupamiento tradicional.

#### 7.4. Etapa 4: Métricas de calidad de agrupamiento

Para evaluar la calidad de los agrupamientos obtenidos, se emplearon como métricas de calidad del agrupamiento, tres índices ampliamente reconocidos en la literatura: **índice de la silueta**, **índice de Calinski-Harabasz** y **índice de Davies-Bouldin** [24].

Estos tres índices proporcionan una valoración complementaria de la estructura del clustering, permitiendo identificar configuraciones óptimas en términos de separación y cohesión de los clusters. Tabla III.

Tabla III  
RESUMEN COMPARATIVO

<b>Índice</b>	<b>Elegir</b>	<b>Evalúa principalmente</b>
Silhouette	Máximo	Separación + Cohesión
Calinski-Harabasz	Máximo	Varianza entre/dentro
Davies-Bouldin	Mínimo	Similitud entre clusters

## 8. RESULTADOS

### 1. *Análisis de los datos:*

Entre los 986 neonatos de la base de datos inicial, todos tenían sospecha de infección pero después de aplicar los criterios en relación a neonatos de 33 o menos semanas con hemocultivos negativos (se excluyeron los neonatos incluidos en las definiciones 1,4 y 5 para duración de antibióticos por sospecha de sepsis inicial con aislamiento de gérmenes bacterianos, virales y por hongos respectivamente) y sin criterios de sepsis clínica (definición 3 aunque hemocultivos negativos, la condición clínica condicionó la persistencia en el uso de antibióticos); se obtuvo un total de 306 neonatos. 180 (59%) de género masculino y 126 (41%) de género femenino cuyas características se muestran en la Tabla IV. para las variables discretas y en la Figura 1. para las variables numéricas.

A los inmaduros extremos (de 25 o menos semanas) correspondieron 86 casos (28%) de los cuales 35 (41%) femeninos y 51 casos (59%) masculinos con una **Mediana:** 0.65 kg, **Rango:** 0.40 kg a 0.96 kg y un **Rango Inter Cuartílico (P25–P75):** 0.57 kg – 0.71 kg para el peso al nacer. Entre 26 a 28 semanas inclusive o pretérminos extremos 90 casos (29%) con 40 femeninos (44%) y masculinos 50 (56%) con datos para el peso al nacer de **Mediana:** 0.88 kg, **Rango:** 0.37 kg a 1.90 kg y **Rango Inter Cuartílico (P25–P75):** 0.68 kg – 1.02 kg. En los pretérminos severos de 29 y 30 semanas se encontraron 42 casos (14%) de los cuales 17 (42%) fueron femeninos y 25 fueron masculinos (58%) con datos de peso al nacer de **Mediana:** 1.35 kg, **Rango:** 0.51 kg a 2.38 kg y **Rango Inter Cuartílico (P25–P75):** 1.16 kg – 1.64 kg. Finalmente, para los llamados pretérminos moderados de 31 a 33 semanas inclusive 88 casos (29%) con 34 femeninos (39%) y 54 casos masculinos (61%) y en cuanto al peso al nacer **Mediana:** 1.81 kg, **Rango:** 0.62 kg a 3.77 kg y **Rango Inter Cuartílico (P25–P75):** 1.45 kg – 2.21 kg.

El estudio por sospecha de infección se inició antes de las 72 horas de vida o lo que corresponde a sepsis temprana en 168 neonatos (55%) y los restantes 138 neonatos (45%) fueron evaluados después de las 72 horas de vida lo que corresponde a sospecha de sepsis tardía. Al momento del estudio en 78 casos (25%) se documentó fiebre, así mismo en 67 casos (22%) se documentó hemorragia intraventricular (no categorizada en la base de datos original). En 19 neonatos (6%) se reportó presencia de catéter arterial umbilical y en 175 (57%) informaron la presencia de catéter venoso central. En los siguientes 28 días de evolución al inicio del estudio de sepsis 157 (51%) neonatos cursaron con ventilación durante más de una semana y 87 neonatos (28%) no requirieron soporte respiratorio invasivo. En cuanto al apoyo inotrópico, 275 neonatos (90%) no requirieron ningún tipo de medicación después del estudio inicial de sepsis.

La Tasa de Uso de Antibióticos de acuerdo con lo reportado en la columna AUR muestra que en 135 neonatos (44.5%) esta fue menor al 5%, en 100 (32.5%) estuvo entre el 6% al 10%, para 44 neonatos (14%) osciló entre 11% al 20% y en los restantes 27 casos (9%) esta fue mayor al 20%.

La mortalidad reportada fue del 7% y en la comparación de características entre los pacientes fallecidos (N = 20) y no fallecidos (N = 286), se observan las siguientes diferencias relevantes: El

uso de catéter venoso central fue significativamente más frecuente en los pacientes fallecidos (90.0%) en comparación con los no fallecidos (54.9%) ( $p = 0.0018$ ), lo que sugiere que los pacientes con mayor gravedad clínica y necesidad de soporte invasivo presentaron mayor mortalidad. Aunque los pacientes fallecidos presentaron una mediana de edad gestacional ligeramente inferior (27.5 vs 28 semanas) y menor peso al nacimiento (0.77 kg vs 0.96 kg), estas diferencias no fueron estadísticamente significativas (Las variables continuas, al no presentar distribución normal, se compararon mediante el test no paramétrico de **Mann-Whitney U**. Se consideraron significativos los valores de  $p < 0.05$ ). Sin embargo, la tendencia podría indicar mayor vulnerabilidad en los extremos de prematuridad. Los fallecidos presentaron significativamente menos días libres de ventilador (mediana 19 días) que los no fallecidos (20 días), además de menor número de días libres de inotrópico (mediana 26.5 días vs 28 días). Aunque la diferencia es sutil en el ventilador, en el uso de **inotrópicos** podría reflejar mayor inestabilidad hemodinámica en el grupo de fallecidos. Los fallecidos presentaron valores significativamente más altos de AUR (mediana 0.11, RIC 0.07–0.18) en comparación con los no fallecidos (mediana 0.05, RIC 0.03–0.08), con un p-valor muy significativo (ver Tabla V). Esto refuerza la asociación previamente identificada entre AUR elevado y mayor riesgo de mortalidad [14]. No se observaron diferencias significativas en cuanto a sexo, uso de catéter arterial umbilical, fiebre, ni comorbilidad IVH/shunt.

Tabla IV  
DESCRIPCIÓN DE LAS VARIABLES CATEGORICAS UTILIZADAS PARA EL  
AGRUPAMIENTO

<b>Variable</b>	<b>Valor</b>	<b>Frecuencia</b>	<b>Porcentaje (%)</b>
Mujer	0	126	41
Hombre	1	180	59
Mortalidad_gener al_en_30_días	No	286	93
Mortalidad_gener al_en_30_días	Si	20	7
Línea venosa central	No	131	43
Línea venosa central	Si	175	57
Línea arterial umbilical	No	287	94
Línea arterial umbilical	Si	19	6
temperatura	No Fiebre	228	75
temperatura	Fiebre	78	25
Comorbilidad_iv h_o_derivación	No	239	78
Comorbilidad_iv h_o_derivación	Si	67	22

y

Tabla V  
FALLECIDOS VS NO FALLECIDOS

Variable	Fallecidos	No Fallecidos	P-valor
<b>sexo</b>	9 (45.0%)	171 (59.8%)	0.2872
<b>Línea venosa central</b>	18 (90.0%)	157 (54.9%)	0.0018
<b>Línea arterial umbilical</b>	2 (10.0%)	17 (5.9%)	0.3574
<b>temp</b>	4 (20.0%)	74 (25.9%)	0.7911
Comorbilidad_ivh_o derivación	5 (25.0%)	62 (21.7%)	0.9461
Edad gestacional al nacer (semanas)	27.50 (25.00–29.25)	28.00 (25.00–31.00)	0.4677
Peso al nacer (kg)	0.77 (0.62–1.32)	0.96 (0.68–1.55)	0.3185
Edad_de inicio_en_dias	4.00 (2.50–7.00)	3.00 (1.00–6.00)	0.1813
Días_libres_de_intubación	19.00 (14.75–23.25)	20.00 (0.00–28.00)	0.8980
Dias_libre_de inotrópicos	26.50 (23.50–28.00)	28.00 (28.00–28.00)	0.0000
<b>aur</b>	0.11 (0.07–0.18)	0.05 (0.03–0.08)	0.0002

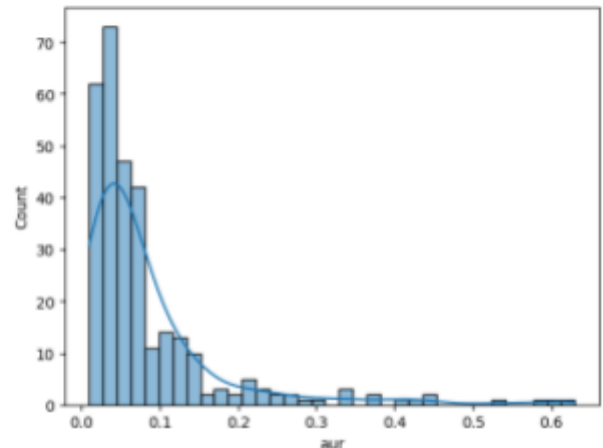
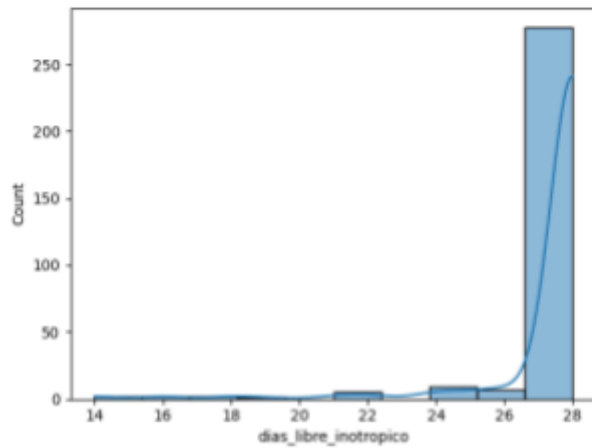
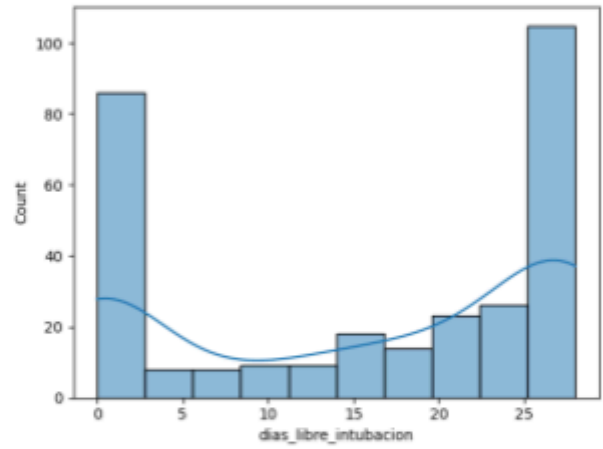
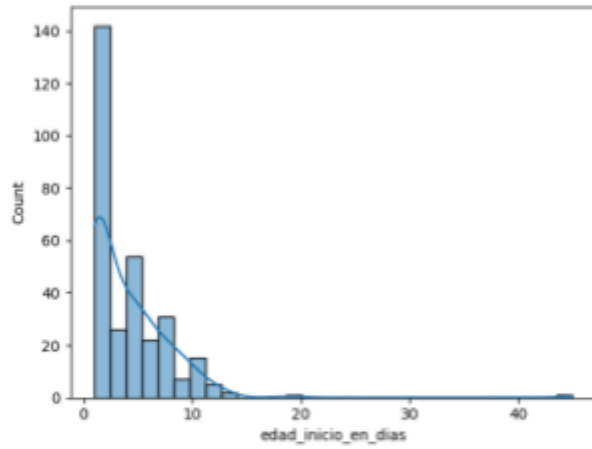
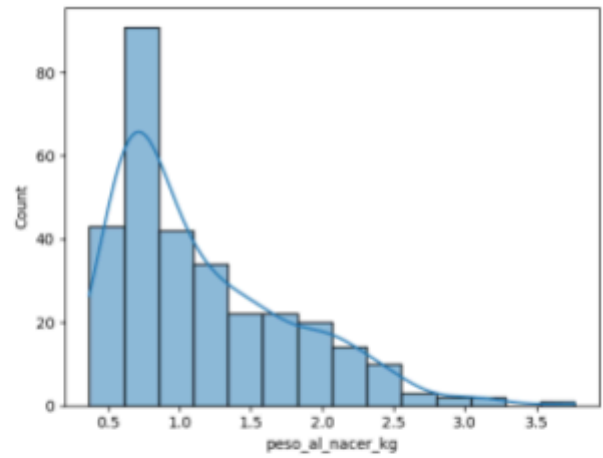
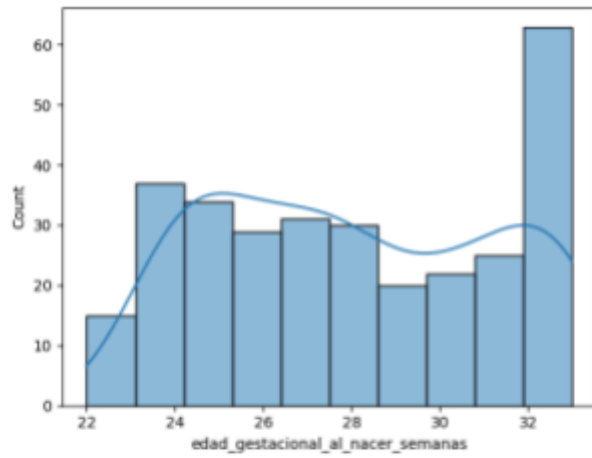


Ilustración SEQ Ilustración \\* ARABIC 1. Variables numéricas

## 2. Modelos de aprendizaje no supervisado:

Para determinar el número óptimo de conglomerados como paso fundamental en los métodos de agrupamiento no supervisados utilizados en el presente análisis se han utilizado varios índices de evaluación cuyos resultados se muestran en la Gráfica. 2 y se resumen en la Tabla VI.

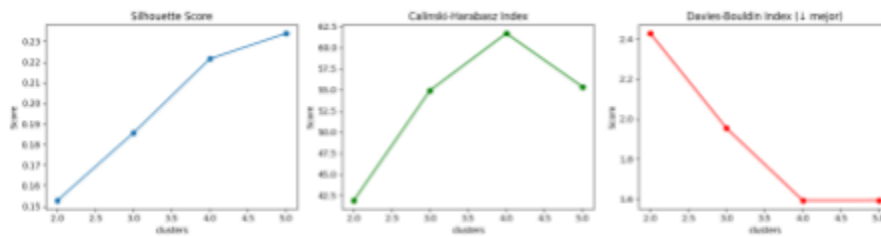


Ilustración SEQ Ilustración \^\* ARABIC 2. Clusters según K Means

KMeans

DBSCAN

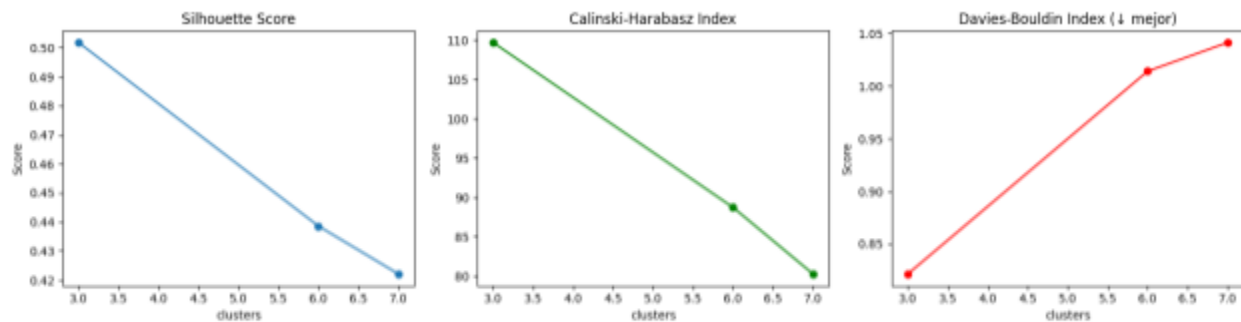
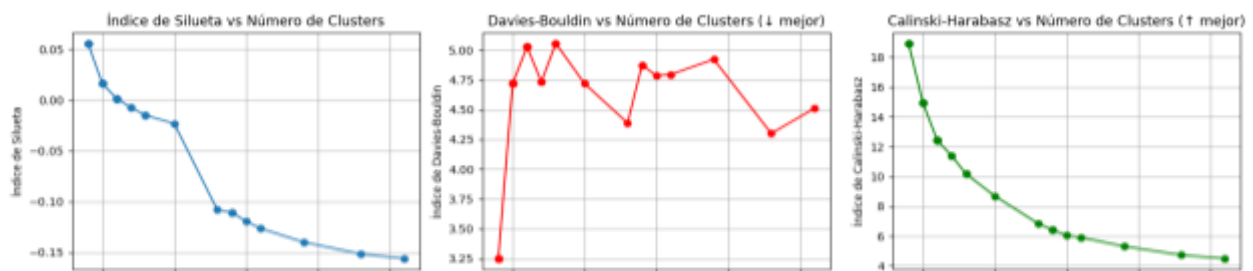


Ilustración SEQ Ilustración \^\* ARABIC 3. Clusters según DBSCAN

JERÁRQUICO



## SOMNET

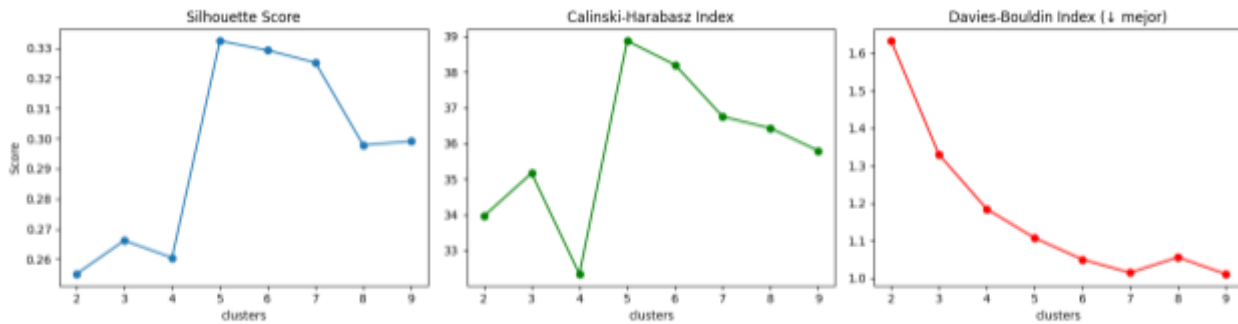


Ilustración SEQ Ilustración \\* ARABIC 5. Clusters según SOMNET

Tabla VI  
CLUSTERS SEGÚN INDICES Y ALGORITMOS

	Índice/ Silueta	Indice Calinski-Harabasz	Indice Davies-Bouldin
KMEANS	5	4	4-5
DBSCAN	3	3	3
JERÁRQUICO	4	4	4
SOMNET	5	5	5-9

Los distintos métodos de agrupamiento mostraron resultados consistentes según los índices de validación interna aplicados. KMeans evidenció un buen desempeño con 4 a 5 clústeres en todos los índices, lo que indica una estructura de grupos sólida y balanceada. DBSCAN fue estable en los tres criterios con 3 clústeres, lo que lo hace útil para detectar agrupaciones densas con formas irregulares. El método jerárquico presentó resultados uniformes con 4 clústeres, lo que sugiere una estructura claramente definida y jerárquica. SOMNet mostró alto rendimiento según los índices de Silueta y Calinski-Harabasz (5 clústeres), aunque el índice de Davies-Bouldin sugirió hasta 9 clústeres, lo que podría reflejar una mayor sensibilidad del modelo a patrones complejos o potencial sobreajuste. En conjunto, KMeans y el método jerárquico ofrecen estructuras robustas y coherentes, mientras que SOMNet puede captar subestructuras más refinadas que requieren validación adicional.

El número de clusters oscila entre 3 y 9 pero utilizando como criterio de tamaño de muestra para escoger el número de clusters y al no haber encontrado un número de referencia de clusters para sepsis neonatal; se escogió 5 para una población de 306 como el indicado teniendo en cuenta que en relación con el criterio de tamaño de muestra cuando esta es pequeña, la capacidad para identificar fenotipos es baja. A diferencia de los estudios supervisados o experimentales, donde el tamaño de muestra se calcula para alcanzar un poder estadístico determinado, los métodos no supervisados

—como clustering o reducción de dimensiones— buscan descubrir patrones o estructuras ocultas en los datos sin etiquetas conocidas, lo que impide establecer de antemano un criterio cuantitativo para determinar el número mínimo de observaciones requeridas. Sin embargo, contar con una muestra suficientemente representativa y diversa es fundamental para lograr resultados estables y generalizables. En estos casos, se recomienda utilizar experimentos de simulación, que permiten estimar la potencia considerando las características reales de los datos. Estudios como el de Koutroulis et al. han demostrado que con simulaciones es posible determinar que una muestra de 150 pacientes permite identificar 4 fenotipos distintos de sepsis en pediatría con una potencia del 80 % [25].

Entre los principales algoritmos de aprendizaje no supervisado —KMeans, DBSCAN, clustering jerárquico y Self-Organizing Maps (SOMNET)— no existe un método universalmente superior, ya que su rendimiento depende del tipo y la estructura de los datos. KMeans es eficiente, rápido y funciona bien cuando los grupos son esféricos y de tamaño similar, siendo adecuado para grandes volúmenes de datos. DBSCAN destaca por identificar agrupaciones de densidad variable y excluir valores atípicos sin necesidad de predefinir el número de clústeres, siendo ideal para conjuntos de datos con ruido. El clustering jerárquico permite explorar relaciones entre subgrupos mediante dendrogramas, aunque puede volverse ineficiente con muestras grandes. SOMNET, por su parte, ha ganado relevancia en análisis de datos de alta dimensionalidad por su capacidad de proyectar estructuras complejas en mapas bidimensionales interpretables. Según la revisión reciente de Raza et al. [26], SOMNET ofrece ventajas notables en visualización y segmentación clínica cuando se requieren interpretaciones multidimensionales no lineales.

En la descripción de las características de cada cluster obtenido con el algoritmo SOMNET Tabla VII, tenemos que en el Cluster 0 (N = 80 pacientes): Este grupo compuesto exclusivamente por pacientes femeninos (100 %) se caracteriza por una ausencia de mortalidad (0 %) y un predominio de pacientes sin fiebre ni comorbilidades neurológicas asociadas (IVH/shunt: 8.8 %). La mediana de edad gestacional fue de 28 semanas (rango: 23-33), con un peso al nacimiento mediano de 0.95 kg. La mediana de edad de inicio fue de 1 día, con 23 días libres de ventilación y 28 días libres de inotrópico. El valor mediano de AUR fue 0.05. Cluster 1 (N = 96 pacientes): Grupo compuesto exclusivamente por pacientes masculinos (100 %), con ausencia de mortalidad. Presentaron alto uso de catéter venoso central (100 %) y fiebre en el 30.2 % de los casos. La mediana de edad gestacional fue 29 semanas (rango: 23-33), peso de 1.15 kg, inicio a los 2 días, con menor cantidad de días libres de ventilación (mediana: 14 días). Los días libres de inotrópico fueron máximos (28 días) y AUR mediana de 0.05. Cluster 2 (N = 44 pacientes): Grupo caracterizado por 100 % de fiebre al inicio, sin mortalidad. Edad gestacional mediana de 27 semanas (rango: 22-33), peso mediano de 0.71 kg. Inicio a los 6 días, con 14 días libres de ventilación, 28 días libres de inotrópico, y AUR bajo (0.04). Bajo uso de catéter venoso central (40.9 %) y sin catéter arterial umbilical. Cluster 3 (N = 66 pacientes): Grupo con predominio masculino (89.4 %), sin mortalidad. No se observaron fiebre ni uso de catéter venoso central. Elevada prevalencia de comorbilidad IVH/shunt (45.5 %). Edad gestacional mediana de 27 semanas (rango: 23-33), peso de 0.91 kg, inicio precoz (2 días). Alta estabilidad ventilatoria (24.5 días libres de ventilador), con 28 días libres de inotrópico y AUR mediana de 0.05. Cluster 4 (N = 20 pacientes): Grupo de alta mortalidad (100 %), con alta frecuencia de catéter venoso central (90 %) y comorbilidades (25 % IVH/shunt). Edad gestacional mediana de 27.5 semanas (rango: 23-33), peso de 0.77 kg. Inicio algo más tardío (4 días), con solo 19 días libres de ventilador, 26.5 días libres de inotrópico. Este grupo mostró el AUR más elevado (mediana: 0.11).

Tabla VII  
CARACTERÍSTICAS DE LOS CLUSTERS

	<b>Cluster 0=80</b>	<b>Cluster 1=96</b>	<b>Cluster 2=44</b>	<b>Cluster 3=66</b>	<b>Cluster 4=20</b>
<b>Sexo</b>	F=80 (100.0%)	M=96 (100.0%)	M=16 (36.4%)	M=59 (89.4%)	M=9 (45.0%)
Mortalidad general en_30_d ías	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	20 (100.0%)
<b>Línea venosa central</b>	43 (53.8%)	96 (100.0%)	18 (40.9%)	0 (0.0%)	18 (90.0%)
<b>Línea arterial umbilical</b>	7 (8.8%)	9 (9.4%)	0 (0.0%)	1 (1.5%)	2 (10.0%)
<b>Temperat ura</b>	0 (0.0%)	29 (30.2%)	44 (100.0%)	1 (1.5%)	4 (20.0%)
Comorbil idad_ivh_ o_derivac ión	7 (8.8%)	17 (17.7%)	8 (18.2%)	30 (45.5%)	5 (25.0%)
Edad gestacion al al nacer (semanas)	23.00 to 33.00 (mediana =28.00)	23.00 to 33.00 (mediana = 29.00)	22.00 to 33.00 (mediana = 27.00)	23.00 to 33.00 (mediana = 27.00)	23.00 to 33.00 (mediana = 27.50)
Peso al nacer (kg)	0.37 to 3.01 (mediana =0.95)	0.45 to 3.20 (mediana = 1.15)	0.40 to 2.69 (mediana = 0.71)	0.48 to 3.77 (mediana = 0.91)	0.48 to 2.27 (mediana = 0.77)
Edad_de inicio_en _días	1.00 to 45.00 (mediana =1.00)	1.00 to 19.00 (mediana = 2.00)	1.00 to 13.00 (mediana = 6.00)	1.00 to 12.00 (mediana = 2.00)	1.00 to 12.00 (mediana = 4.00)
Días_libre s_de_intu bación	0.00 to 28.00	0.00 to 28.00	0.00 to 28.00	0.00 to 28.00	8.00 to 27.00

	(mediana = 23.00)	(mediana = 14.00)	(mediana = 14.00)	(mediana = 24.50)	(mediana = 19.00)
Dias_libre_de_inotrópicos	22.00 to 28.00 (mediana = 28.00)	14.00 to 28.00 (mediana = 28.00)	22.00 to 28.00 (mediana = 28.00)	28.00 to 28.00 (mediana = 28.00)	18.00 to 28.00 (mediana = 26.50)
<b>AUR</b>	0.01 to 0.42 (mediana = 0.05)	0.01 to 0.61 (mediana = 0.05)	0.01 to 0.44 (mediana = 0.04)	0.01 to 0.53 (mediana = 0.05)	0.02 to 0.63 <b>(mediana = 0.11)</b>

Finalmente se evaluó la asociación entre el valor de AUR y la mortalidad a 30 días, así como las diferencias en AUR entre los diferentes clusters identificados. El análisis no paramétrico de Kruskal-Wallis mostró que el valor de AUR difirió significativamente entre los clusters ( $H = 17.98$ ,  $p = 0.0012$ ), siendo notablemente más elevado en el cluster 4. Adicionalmente, la comparación de AUR entre los pacientes fallecidos y los supervivientes mediante la prueba de Mann-Whitney U reveló que los pacientes que fallecieron presentaron valores de AUR significativamente mayores que aquellos que sobrevivieron ( $U = 4258.50$ ,  $p = 0.0002$ ).

La caracterización de los *clusters* contó con el respaldo de cuatro pediatras neonatólogos vinculados a diferentes entornos clínicos. Este panel de expertos evaluó las variables distintivas de cada agrupación con el objetivo de definir y validar los atributos predominantes de cada segmento, asegurando una interpretación alineada con la práctica médica actual. Tras el análisis y validación por parte del panel de expertos, se establecieron las categorías clínicas para cada uno de los grupos identificados. El **clúster 0** fue definido como 'prematureo sin morbilidad específica'. Por su parte, los grupos restantes fueron clasificados según la patología predominante: el **clúster 1** se vinculó a afecciones gastrointestinales; el **clúster 2**, a complicaciones respiratorias; el **clúster 3**, a sintomatología neurológica; y, finalmente, el **clúster 4** fue categorizado bajo el cuadro de respuesta inflamatoria sistémica sin germen aislado.

## 9. DISCUSIÓN

La sepsis neonatal carece de una definición internacional consensuada. En la literatura, también hay inconsistencias sobre el momento de iniciar y suspender antibióticos y sobre el monitoreo según riesgo materno, antibióticos intraparto, prematuridad y presentación clínica. Ante el temor a desenlaces graves, se usa tratamiento precoz, amplio y prolongado, con posibles efectos adversos sobre todo sin evidencia microbiológica como alteración del microbioma, enterocolitis, infecciones fúngicas, mortalidad y resistencia microbiana [15].; aunque varias guías proponen suspender a las 36–48 h si hay mejoría, hemocultivos negativos y laboratorio normal, la evidencia de alta calidad es escasa.

La evidencia disponible presenta calidad dispar, heterogeneidad clínica y pocos ensayos comparables entre sí, lo que limita conclusiones; de otro lado, agrupar estrategias basadas en biomarcadores puede aumentar la heterogeneidad. Se requieren ensayos clínicos aleatorizados rigurosos y transparentes, metanálisis en red cuando haya estudios más homogéneos y evaluaciones económicas para orientar guías costo-efectivas [28] .

Los ensayos controlados aleatorizados siguen siendo el estándar de oro para evaluar la eficacia de las intervenciones médicas, sentando las bases de la medicina basada en la evidencia. Al asignar aleatoriamente a los participantes a grupos de tratamiento o control, se busca minimizar el sesgo y establecer la causalidad. Sin embargo, en cuidados intensivos con frecuencia estos ensayos no han logrado demostrar beneficios significativos. Esta falta de resultados positivos no significa necesariamente que las intervenciones en estudio sean ineficaces; más bien, pone de relieve la complejidad y heterogeneidad de los pacientes críticos. Los pacientes de la UCI presentan una heterogeneidad sustancial.

Estas variaciones dificultan la detección de efectos significativos del tratamiento, ya que en los ensayos clínicos aleatorizados tradicionales se mide principalmente el efecto promedio del tratamiento, que puede enmascarar diferencias importantes en las respuestas de los pacientes, lo que podría ocultar los beneficios para algunos subgrupos y pasar por alto los daños en otros. A esta variabilidad se le llama efectos heterogéneos del tratamiento. Al analizarse como una sola población en un ensayo clínico aleatorizado, estos efectos opuestos pueden anularse mutuamente, dando como resultado un resultado general neutral o no concluyente [29].

Abordar este desafío requiere un enfoque más refinado para la selección y el análisis de pacientes en ensayos clínicos. En lugar de asumir una población de pacientes según los criterios de elegibilidad y efectos del tratamiento uniformes en todos los pacientes de la UCI, la identificación de subgrupos más homogéneos con perfiles de riesgo, genéticos, metabólicos, fisiológicos, de comorbilidades y de gravedad de la enfermedad similares podría mejorar la capacidad de detectar efectos significativos del tratamiento.

La inteligencia artificial y el aprendizaje automático ofrecen herramientas poderosas para enriquecer los ensayos clínicos al identificar subgrupos relevantes de pacientes y reducir la heterogeneidad al descubrir los efectos individualizados del tratamiento. Las técnicas de inteligencia artificial que están diseñadas para reducir eficazmente la heterogeneidad en los ensayos clínicos, mejorando el diseño de los ensayos,

optimizando la selección de pacientes y permitiendo la toma de decisiones en tiempo real en cuidados intensivos incluyen el agrupamiento, el modelado predictivo, la inteligencia artificial causal, el aprendizaje por refuerzo y la inteligencia artificial generativa .

El agrupamiento es una técnica de aprendizaje no supervisado en el aprendizaje automático que reúne datos similares en función de características compartidas. A diferencia de los análisis de subgrupos tradicionales, que se basan en variables predefinidas, los algoritmos de agrupamiento pueden descubrir patrones ocultos en conjuntos de datos complejos al considerar múltiples factores simultáneamente. En cuidados críticos, la agrupación puede ayudar a definir subpoblaciones de pacientes más homogéneas para los ensayos clínicos aleatorizados. Al adaptar las intervenciones a estos grupos específicos, los investigadores pueden enriquecer los ensayos, reduciendo la variabilidad y aumentando la probabilidad de detectar los verdaderos efectos del tratamiento. Además, la agrupación puede aplicarse posteriormente a los ensayos completados, revelando una heterogeneidad previamente no reconocida [30].

El análisis de aprendizaje automático no supervisado derivó cinco fenotipos computables con cumplimiento de tres de los cinco criterios de "camino a seguir" [31] que demuestran relevancia clínica con diferencias en el sexo, las características al momento del estudio inicial de sepsis en relación con fiebre, con presencia de hemorragia intraventricular, en el uso de catéter venoso central, la necesidad de soporte ventilatorio y de soporte inotrópico posterior al inicio del estudio; pero sobre todo en la Tasa del Uso de Antibióticos . La relación entre los fenotipos computables derivados y la mortalidad demostró un incremento en la probabilidad de mortalidad con el mayor uso de antibióticos con hemocultivos negativos. Hay limitaciones en el presente análisis, como la naturaleza retrospectiva del estudio, el uso de una base de datos externa y la ausencia de variables clínicas más detalladas, especialmente aquellas relacionadas con biomarcadores o hallazgos diagnósticos específicos, que habrían permitido una caracterización más precisa de los cuadros clínicos. Esta falta de información impidió clasificar con certeza los clústeres como respiratorios, cardiovasculares, gastrointestinales, neurológicos o multisistémicos. Además, al tratarse de una cohorte de un solo centro, la generalización de los hallazgos requiere cautela.

Desde el punto de vista metodológico, los algoritmos de aprendizaje no supervisado empleados (KMeans, DBSCAN, clustering jerárquico y SOMNet) tienen limitaciones inherentes: pueden verse afectados por la escala de las variables, la presencia de ruido, la selección del número de clústeres y la sensibilidad a valores atípicos. En especial, algoritmos como KMeans suponen que los grupos son esféricos y de tamaño similar, lo que no siempre refleja la realidad clínica. DBSCAN puede no funcionar bien cuando las densidades de los clústeres varían, y SOMNet requiere interpretaciones cuidadosas por la complejidad de sus representaciones. En este contexto, aunque estos modelos permiten descubrir patrones latentes útiles para la estratificación del riesgo, su utilidad clínica dependerá de su validación externa, reproducibilidad y traducción a herramientas interpretables para el personal clínico, con una integración adecuada en flujos de trabajo que garanticen decisiones basadas en evidencia y no sólo en resultados computacionales.

Sin embargo, en un estudio de reciente publicación [32] en donde se utilizaron herramientas del aprendizaje no supervisado para agrupamiento en pacientes con sepsis neonatal los autores identificaron tres grupos robustos que denominaron endotipos, dada su singularidad fisiopatológica, con base en la expresión de 100 genes relacionados con mortalidad. Los pacientes asignados al endotipo A presentaron probabilidades ajustadas de mortalidad significativamente mayores (22 % frente a 0 %,  $p = 0.003$ ). Cabe

destacar que los neonatos pertenecientes al endotipo A presentaron además tasas más bajas de sepsis de inicio temprano, tasas de identificación de patógenos más bajas que aquellos con endotipos B y C, y mayor disfunción cardíaca (61 % frente a 31 %,  $p = 0.025$ ) como en nuestro fenotipo 4 que cursó con mayor mortalidad, cultivos negativos y mayor uso de soporte inotrópico en relación con disfunción cardíaca.

Los hallazgos revelaron una respuesta inmunitaria innata hiperactiva impulsada por los progenitores de neutrófilos y la represión de la señalización de los linfocitos T en los endotipos perjudiciales, lo que sugiere que la desregulación inmunitaria, y no la edad gestacional ni la detección de patógenos, es el principal determinante de los resultados de la sepsis neonatal. Los autores concluyeron que el perfil de expresión génica se puede utilizar para desentrañar la heterogeneidad de la sepsis neonatal y, a la espera de una mayor validación, la subclasificación basada en la expresión génica se puede utilizar para identificar a los neonatos en riesgo e informar la selección de terapias específicas para la sepsis . Hallazgos que coinciden con los reportados por otro estudio también de reciente publicación en el que se aplicó la técnica de aprendizaje no supervisado tipo Algoritmo Jerárquico para definir los fenotipos del perfil genético en neonatos con sepsis neonatal[33].

## 10. RECOMENDACIONES Y TRABAJOS FUTUROS

Con base en los hallazgos del presente estudio, se recomienda fortalecer la toma de decisiones clínicas mediante el uso de herramientas basadas en aprendizaje automático dentro de las Unidades de Cuidado Intensivo Neonatal (UCIN), como apoyo a la valoración médica tradicional, especialmente en neonatos con hemocultivos negativos. Asimismo, es fundamental estandarizar y mejorar la calidad de los registros clínicos, incorporando variables como biomarcadores, puntuaciones de severidad y evolución fisiológica, lo que permitiría construir modelos predictivos más robustos y clínicamente útiles. Para reducir los riesgos asociados al uso prolongado de antibióticos, se sugiere implementar protocolos institucionales que promuevan la reevaluación periódica del tratamiento empírico en ausencia de confirmación microbiológica. Finalmente, se destaca la importancia de fomentar el trabajo interdisciplinario entre profesionales de la salud, ingeniería biomédica y ciencias de datos, para asegurar una integración efectiva y responsable de estas tecnologías en la práctica clínica neonatal.

En cuanto a los trabajos futuros, será prioritario validar los fenotipos clínicos identificados en esta investigación mediante su aplicación en bases de datos externas de diversos entornos hospitalarios, lo cual permitirá evaluar su generalización y utilidad clínica. Como siguiente paso metodológico, se plantea el desarrollo de modelos supervisados que utilicen estos fenotipos como etiquetas predictivas para anticipar el riesgo de mala evolución clínica. También se recomienda incorporar nuevas variables, como biomarcadores, hallazgos de imagen y datos longitudinales (por ejemplo, la evolución de signos vitales), que podrían enriquecer los modelos no supervisados actuales. A partir de los resultados obtenidos, se propone el diseño de una herramienta de apoyo a decisiones clínicas basada en inteligencia artificial, integrada a los sistemas de historia clínica electrónica. Finalmente, futuras investigaciones deberían evaluar el impacto clínico y económico de estas estrategias, particularmente en términos de reducción de morbilidad, mortalidad, estancia hospitalaria y costos asociados, para sustentar su implementación a nivel institucional.

## 11. CONCLUSIÓN

Este estudio aplicó múltiples algoritmos de aprendizaje automático no supervisado —incluyendo KMeans, DBSCAN, clustering jerárquico y SOMNet— para analizar patrones clínicos en neonatos menores de 33 semanas con sospecha de sepsis y hemocultivos negativos. La combinación de métodos permitió identificar cinco fenotipos computacionales distintos, con el algoritmo SOMNet mostrando una alta capacidad para representar estructuras complejas de los datos clínicos. Estos clústeres emergieron a partir de variables clínicas comunes como edad gestacional, peso al nacer, días libres de soporte respiratorio e inotrópico, y valores de AUR, lo que demuestra el valor del análisis no supervisado en contextos donde no existen etiquetas diagnósticas claras. El fenotipo identificado como cluster 4, asociado con mayor mortalidad, resalta la posibilidad de que algunos casos presenten una etiología no infecciosa, lo que podría modificar el enfoque clínico y evitar tratamientos antibióticos innecesarios. Este tipo de análisis permite a los clínicos ir más allá del juicio individual y considerar patrones latentes en los datos que podrían anticipar riesgo clínico. Si se valida en cohortes externas, esta metodología podría convertirse en una herramienta de apoyo para la estratificación del riesgo, la individualización del tratamiento y la optimización del uso de antibióticos en la práctica neonatal.

## 12. REFERENCIAS

- [1] Sanchez-Pinto LN, *et al.* Development and validation of the Phoenix criteria for pediatric sepsis and shock septic. JAMA published online January 21, 2024 doi: 10.1001/jama.2024.0196.
- [2] Depani SJ, Ladhani S, Heath PT, Lamagni TL, Johnson AP, Pebody RG, *et al.* The contribution of infections to neonatal deaths in England and Wales. *Pediatr Infect Dis J.* 2011;30(4):345–7.
- [3] Hayes, R. *et al.* Neonatal sepsis definitions from randomised clinical trials. *Pediatr. Res.* 2023; 93:1141-1148.
- [4] Clark RH, Bloom BT, Spitzer AR, Gerstmann DR. Empiric use of ampicillin and cefotaxime, compared with ampicillin and gentamicin, for neonates at risk for sepsis is associated with an increased risk of neonatal death. *Pediatrics.* 2006;117(1):67–74.
- [5] Tripathi N, Cotton CM, Smith P. Antibiotic use and misuse in the neonatal intensive care unit. *Clin Perinatol.* 2012;39(1):61–8.
- [6] Russell, N., Barday, M., Dramowski, A., Sharland, M., Bekker, A. Early-versus Late-Onset Sepsis in Neonates—Time to Shift the Paradigm? *Clin. Microbiol. Infect.* **2023**, 30, 38–43.
- [7] Russell AB, Sharland M, Heath PT. Improving antibiotic prescribing in neonatal units: time to act. *Arch Dis Fetal Neonatal.* 2012;97:F141–6.
- [8] Cotten CM, Taylor S, Stoll B, Goldberg RN, Hansen NI, Sánchez PJ, *et al.* Prolonged duration of initial empirical antibiotic treatment is associated with increased rates of necrotizing enterocolitis and death for extremely low birth weight infants. *Pediatrics.* 2009;123:58–66.
- [9] Blackburn RM, Verlander N, Heath P, Muller-P B. The changing antibiotic susceptibility of bloodstream infections in the first month of life: informing antibiotic policies for early- and late-onset neonatal sepsis. *Epidemiol Infect.* 2014;142(4):803–11.
- [10] Geneva: World Health Organization. Ten threats to global health in 2019. <https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019> [Accessed 15 March 2024].
- [11] World Health Assembly 68. Sixty-eighth World Health Assembly, Geneva, 18-26 May 2015: Agenda Item 15.1 Global action plan on antimicrobial resistance. ([http://apps.who.int/gb/ebwha/pdf\\_files/WHA68/A68\\_R7-en.pdf](http://apps.who.int/gb/ebwha/pdf_files/WHA68/A68_R7-en.pdf)) [Accessed 15 March 2024].
- [12] Flannery DD, Zevallos Barboza A, Mukhopadhyay S, Gerber JS, McDonough M, Shu D, Hennessy S, Wade KC, Puopolo KM. Antibiotic use among extremely low birth-weight infants from 2009 to 2021: a retrospective observational study. *Arch Dis Child Fetal Neonatal Ed.* 2024 Jul 20:fetalneonatal-2023-326734.
- [13] O’Sullivan B. *et al.* Machine learning applications on neonatal sepsis treatment: a scoping review *BMC Infectious Diseases* 2023; 23:441-451.
- [14] Benoni R, Balestri E, Endrias T, Tolera J, Borellini M, Calia M, Biasci F, Pisani L. Exploring the use of cluster analysis to assess antibiotic stewardship in critically-ill neonates in a low resource setting. *Antimicrob Resist Infect Control.* 2023 Oct 31;12(1):119. doi: 10.1186/s13756-023-01325-w.
- [15] Ting JY, Synnes A, Roberts A, *et al.* Association between antibiotic use and neonatal mortality and morbidities in very low-birth-weight infants without culture-proven sepsis or necrotizing enterocolitis. *JAMA Pediatr* 2016;170(12):1181–7.
- [16] Sullivan BA, Grundmeier RW. Machine Learning Models as Early Warning Systems for Neonatal Infection. *Clin Perinatol.* 2025 Mar;52(1):167-183. doi: 10.1016/j.clp.2024.10.011. Epub 2024 Nov 27.
- [17] Lawrence SM, Wynn JL, Kimberlin DW, Cantey JB. Investigating antibiotics in the NICU and patient safety. *Front Cell Infect Microbiol.* 2025 May 1;15:1563940. doi: 10.3389/fcimb.2025.1563940.
- [18] Ting JY, Roberts A, Sherlock R, *et al.* Duration of initial empirical antibiotic therapy and outcomes in very low birth weight infants. *Pediatrics* 2019;143(3).
- [19] Ting JY, Synnes A, Roberts A, *et al.* Association of antibiotic utilization and neuro-developmental outcomes among extremely low gestational age neonates without proven sepsis or necrotizing enterocolitis. *Am J Perinatol* 2018;35(10):972–8.
- [20] Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques* (3rd ed.). Elsevier.
- [21] Ostapenko S, Schmatz M, Srinivasan L *et al.* . Neonatal sepsis registry: Time to antibiotic dataset. *Data in brief.* 2019 104788.
- [22] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., and Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- [23] Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8), 651–666. <https://doi.org/10.1016/j.patrec.2009.09.011>
- [24] Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)*, 226–231.

- [25] Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: A review. *ACM Computing Surveys (CSUR)*, 31(3), 264–323. <https://doi.org/10.1145/331499.331504>
- [26] Gibson, S., van Vreeswijk, C., & Hennequin, G. Self-organizing maps for interpretable dimensionality reduction and clustering of high-dimensional data. *Neural Networks*, (2023) 160, 155–168.
- [27] Vinué, L., Pons, J., & Martí, R.. A comprehensive review of internal clustering validation measures: Addressing issues in Silhouette, Calinski–Harabasz, Davies–Bouldin indices. *Pattern Recognition Letters*, (2023) 171, 83–94.
- [28] Koutroulis I, Velez T, Wang T, Yohannes S, Galarraga JE, Morales JA, et al. Pediatric sepsis phenotypes for enhanced therapeutics: an application of clustering to electronic health records. *J Am Coll Emerg Physicians Open*. (2022) 3(1):e12660. doi: 10.1002/emp2.12660
- [29] A. Guérin, P. Chauvet, and F. Saubion, “A Survey on Recent Advances in Self-Organizing Maps,” *arXiv preprint*, arXiv:2501.08416, 2024. [Online]. Available: <https://arxiv.org/abs/2501.08416>
- [30] Fung GP, Ting JY. Re-Thinking the Norms of Antibiotic Prescribing in the Neonatal Intensive Care Unit. *Clin Perinatol*. 2025 Mar;52(1):133-146. doi: 10.1016/j.clp.2024.10.009. Epub 2024 Nov 29. PMID: 39892948.
- [31] Feng K, Zhang T, Hua Z. Discontinuation of empirical antibiotics in suspected neonatal early-onset sepsis: a systematic review and meta-analysis. *Pediatr Res*. 2025 Jul 29. doi: 10.1038/s41390-025-04290-9. PMID: 40731093.
- [32] Oh W, Veshtaj M, Sakhuja A. Enriching patient populations in ICU trials: reducing heterogeneity through machine learning. *Curr Opin Crit Care*. 2025 Aug 1;31(4):410-416.
- [33] Feuerriegel S, Frauen D, Melnychuk V, Schweisthal J, Hess K, Curth A, Bauer S, Kilbertus N, Kohane IS, van der Schaar M. Causal machine learning for predicting treatment outcomes. *Nat Med*. 2024 Apr;30(4):958-968. doi: 10.1038/s41591-024-02902-1. Epub 2024 Apr 19. PMID: 38641741.
- [34] DeMerle KM, Angus DC, Baillie JK, et al. Sepsis subclasses: a framework for development and interpretation. *Crit Care Med*. 2021;49:748–59.
- [35] Al Gharaibeh FN, Huang M, Wynn JL, Kamaleswaran R, Atreya MR. Transcriptomic mortality signature defines high-risk neonatal sepsis endotype. *Front Immunol*. 2025 Jun 27;16:1601316. doi: 10.3389/fimmu.2025.1601316. PMID: 40655143; PMCID: PMC12245816.
- [36] Wang H, Yang R, Chen N, Li X. Heterogeneity of Neutrophils and Immunological Function in Neonatal Sepsis: Analysis of Molecular Subtypes Based on Hypoxia-Glycolysis-Lactylation. *Mediators Inflamm*. 2025 Mar 26;2025:5790261. doi: 10.1155/mi/5790261. PMID: 40177399; PMCID: PMC11964727.
- [37] Gerard R, Dewitte A, Gross F, Pradeu T, Lemoine M, Goret J, Mamani-Matsuda M. Is "pre-sepsis" the new sepsis? A narrative review. *PLoS Pathog*. 2025 Jul 31;21(7):e1013372.

