

# ANÁLISIS ESTADÍSTICO DE LAS PRINCIPALES CAUSAS DE MUERTE A NIVEL MUNDIAL DESDE 1990 A 2022 UTILIZANDO TÉCNICAS MULTIVARIANTES

## STATISTICAL ANALYSIS OF THE LEADING CAUSES OF DEATH WORLDWIDE FROM 1990 TO 2022 USING MULTIVARIATE TECHNIQUES

Dennys Daquilema<sup>1</sup>, Lidia Castro<sup>2</sup>

{dennys.daquilema@unach.edu.ec<sup>1</sup>, lidia.castro@unach.edu.ec<sup>2</sup>}

Fecha de recepción: 30/05/2025 / Fecha de aceptación: 12/06/2025 / Fecha de publicación: 15/06/2025

**RESUMEN:** El presente estudio analiza la evolución de las principales causas de muerte a escala mundial y nacional entre 1990 y 2021, y genera proyecciones para el año 2022 mediante técnicas estadísticas multivariantes aplicadas a datos del Global Burden of Disease (GBD). La investigación parte del reconocimiento de que los enfoques univariados tradicionales son limitados para captar la complejidad estructural de la mortalidad, por lo que se adoptó un enfoque mixto que integra análisis exploratorio, Análisis de Componentes Principales (ACP) y algoritmos de predicción como redes neuronales LSTM y Random Forest. La metodología incluyó la depuración y estandarización de los datos mediante tasas ajustadas por cada 100 000 habitantes y puntajes Z, lo que permitió comparaciones válidas entre países y años. El ACP redujo 32 causas de muerte a cinco dimensiones epidemiológicas latentes, que facilitaron la identificación de patrones regionales. Se evidenció que, en países de bajos ingresos, prevalecen enfermedades transmisibles, mientras que en los de altos ingresos, como Estados Unidos, predominan enfermedades crónicas y trastornos mentales. El modelo Random Forest mostró el mejor rendimiento ( $R^2 = 0.99$ ; MAE = 512), superando a LSTM, y permitió estimar que en 2022 las enfermedades cardiovasculares serían la principal causa global de defunción, con un 35,6 % del total. Las variables predictoras incluyeron el país, el año y la cantidad de muertes por cada enfermedad, mientras que la variable respuesta fue la tasa ajustada de mortalidad por causa. En conjunto, los resultados validan la utilidad de la estadística multivariante para el análisis de fenómenos complejos en salud pública y para la orientación de políticas basadas en evidencia.

**Palabras clave:** *Técnicas Multivariantes, Random Forest, ACP, tasa de mortalidad, predicción estadística, Redes Neuronales LSTM, causas de muerte, machine learning*

**ABSTRACT:** This study analyzes the evolution of the leading causes of death at both global and national levels from 1990 to 2021, and generates projections for 2022 using multivariate statistical techniques applied to data from the Global Burden of Disease (GBD). Recognizing the limitations of traditional univariate approaches to capture the structural complexity of mortality, the research adopts a mixed-methods strategy that integrates exploratory analysis, Principal Component Analysis (PCA), and predictive algorithms such as Long Short-Term Memory (LSTM) neural networks and Random Forest. The methodology included data

<sup>1</sup>Universidad Nacional de Chimborazo, <https://orcid.org/0000-0001-9830-0471>.

<sup>2</sup>Universidad Nacional de Chimborazo, <https://orcid.org/0000-0002-0471-2879>.

cleaning and standardization through age-adjusted rates per 100,000 inhabitants and Z-scores, enabling valid comparisons across countries and time. PCA reduced 32 causes of death to five latent epidemiological dimensions, which facilitated the identification of regional patterns. Results showed that in low-income countries, transmissible diseases are more prevalent, whereas in high-income countries such as the United States, mortality is primarily driven by chronic diseases and mental disorders. The Random Forest model demonstrated superior performance ( $R^2 = 0.99$ ; MAE = 512), outperforming LSTM, and allowed for estimating that in 2022, cardiovascular diseases would account for 35.6% of global deaths. Predictor variables included country, year, and the number of deaths per cause, while the response variable was the adjusted mortality rate by cause. Overall, the findings support the use of multivariate statistics as a robust tool for analyzing complex public health phenomena and informing evidence-based policy-making.

**Keywords:** *Multivariate Techniques, Random Forest, PCA, mortal rate, Statistical Prediction, LSTM Neural Networks, Causes of Death*

## INTRODUCCIÓN

El estudio de las causas de muerte no se limita únicamente al ámbito médico, sino que trasciende hacia la construcción de métricas que cuantifican el funcionamiento de los sistemas de salud, mide la eficacia de las políticas públicas, y que se relaciona con el nivel de progreso económico y social de un país por medio de la esperanza de vida (1). En años actuales, esos enfoques han sido entendidos, llevando al estudio de las causas de muerte a ser uno de los aspectos más estudiados, hasta convertirse en insumos para la toma de decisiones en planeación sanitaria, evaluación de planes médicos preventivos y de proyecciones demográficas, como lo reconoce la Organización Mundial de la Salud (2).

A pesar de su importancia, gran parte de los estudios existentes se centran únicamente en estudios univariados, y a enfoques netamente descriptivos que, aunque son adecuados para generar visualizaciones, establecer rankings simples, o detectar aumentos o disminuciones en enfermedades específicas, presentan debilidades y limitaciones en estructurar relaciones complejas entre variables, modelos de comportamiento, y evoluciones temporales. Como solución de estas carencias, la estadística multivariante ofrece un conjunto de técnicas capaces de abordar simultáneamente múltiples dimensiones, extrayendo comportamientos y patrones ocultos, constituyéndose en la mejor manera de mantener y comprender la información que los datos poseen (3).

La aplicación de estas técnicas multivariantes requiere bases de datos robustas, como la proporcionada por la Global Burden of Disease (GBD), que es elaborada por un centro de investigaciones afiliado a la Universidad de Washington denominado Institute for Health Metrics and Evaluation (IHME). Esta base de datos se constituye en uno de los mayores esfuerzos que busca recopilar, estandarizar y almacenar la información de las causas de muerte a nivel mundial. El GBD integra los datos provenientes de los registros civiles, partidas médicas, y todo tipo de registros de defunción de más de 250 países y territorios. Para su levantamiento, la IHME cuenta con procedimientos rigurosos de corrección de registros, imputación de valores faltantes, ajuste por edad y género, así como el empleo de modelos bayesianos jerárquicos que buscan reducir el sesgo sistemático, constituyéndola en

una de las fuentes más fiables en el ámbito de datos de mortalidad (4).

En lo concerniente a trabajos relacionados, diversas metodologías se han utilizado para estudiar las causas de muerte, desde técnicas estadísticas tradicionales hasta enfoques avanzados basados en inteligencia artificial, por ejemplo, con el propósito de superar limitaciones asociadas a la codificación manual, en 2011 se desarrolló el software IRIS, el cual automatiza la codificación de causas de defunción con una precisión superior al 80% (5). Estudios más recientes han mejorado esta precisión utilizando redes neuronales (6) o convolucionales (7), alcanzando métricas de hasta el 98.44% de exactitud.

La data GBD ha sido base de múltiples investigaciones, por ejemplo, un análisis proyectivo sobre depresión y ansiedad estimó una reducción hacia 2044, con mayores tasas históricas en mujeres y adultos mayores (8). En otro caso, se evaluó la hipertensión arterial pulmonar, encontrando una reducción global y diferencias de mortalidad por sexo (9). El estudio realizado por Bishop en (10), parte de la iniciativa que busca refutar el enfoque tradicional que atribuye la mortalidad a una única causa, para lo cual emplearon clustering y técnicas any-mention. En cuanto a enfermedades específicas, se han empleado modelos ARIMA para estudiar esquistosomiasis, observando una tendencia decreciente y un error de predicción inferior al 4% (11). Otro trabajo analizó el vínculo entre el IMC y 282 causas de muerte, destacando un rol de enfermedades cardiovasculares y renales, así como el uso de regresión multivariable y tasas ajustadas por edad (12).

En el campo de la reducción dimensional, técnicas como ACP y la Transformada Discreta de Wavelet (DWT) han sido comparadas, obteniendo esta última una precisión del 99% y una velocidad de ejecución notable (13). Finalmente, el Índice de Calidad de Atención (QCI), construido mediante ACP, ha sido utilizado para evaluar servicios odontológicos (14), y muertes por cáncer ginecológico (15), revelando brechas según edad y nivel socioeconómico.

Estos antecedentes refuerzan la utilidad de combinar análisis multivariantes con modelos de predicción para abordar la complejidad inherente a la carga global de enfermedades, justificando la aplicación metodológica del presente estudio, y cuya naturaleza radica en la necesidad de comprender cómo se distribuyen y evolucionan las principales causas de muerte a escala mundial (16).

El alcance de este análisis abarca el período de 1990 a 2022, empleando datos del Global Burden of Disease (GBD) de 1990 a 2021 para más de 190 países, e integrando técnicas como el Análisis de Componentes Principales (ACP) y algoritmos de predicción como Random Forest y redes neuronales LSTM para hacer predicciones para el año 2022 (17).

El objetivo general del estudio es aplicar técnicas multivariantes para analizar las principales razones de muerte a nivel mundial y generar proyecciones para 2022. Como objetivos específicos, se plantea: i) depurar la base de datos mediante técnicas estadísticas ii) agrupar enfermedades mediante el uso de técnicas multivariantes iii) construir modelos predictivos, y iv) evidenciar las principales causas de muerte v) identificar el comportamiento de las variables a lo largo del tiempo, y determinar las 10 principales causas de muerte. Este enfoque busca proporcionar una base empírica robusta para el diseño de políticas de salud

pública sustentadas en evidencia analítica (18).

## MATERIALES Y MÉTODOS

El presente estudio es de tipo cuantitativo, observacional, y longitudinal, y utilizó la base de datos GBD proporcionada por el IHME, que levanta los datos de las principales causas de muerte a nivel mundial en un periodo correspondiente de 1990-2021. Los datos fueron obtenidos a partir del sitio web (19), y cuantifica la información de más de 250 países y territorios de las 32 principales causas de muerte, de la cual se excluyeron los datos de COVID-19 con el objetivo de reducir sesgos temporales. Dado que los datos reales no contemplan información del año 2022, se realizaron proyecciones estadísticas para este año.

Tras completar el proceso de depuración de la data mediante herramientas de Rstudio, se procedió a estandarizar las variables con la tasa por 100 000 habitantes que será empleada para ciertas visualizaciones exploratorias, mientras que para los análisis se normalizó mediante Z-score. El análisis exploratorio inicia con la realización de diagramas de cajas y bigotes (boxplot) de todos los países, en cada gráfico se representan las 32 causas de muerte y sus dispersiones, mientras que con los raincloud plots se mostró la distribución, densidad y variabilidad de cada enfermedad, para luego verificar su tipo de distribución mediante los QQ-plots y la prueba de Lilliefors. Mediante la herramienta geom\_spatvector se representó la distribución de población y la cantidad de muertes totales en 2021, que es el dato más actual de la base de datos. La matriz de correlaciones basada en los datos estandarizados cuantificó la asociación entre las causas de muerte, para finalmente pasar a la representación de las series de tiempo de cada enfermedad.

Para reducir la dimensión del conjunto de observaciones, se aplicó un análisis de componentes principales (ACP) basado en la matriz de correlaciones mediante la función PCA de R. El criterio de Kaiser determinó el número de dimensiones a ser consideradas, y en base a estas se agruparon las variables en cinco grupos que acumulan la mayor parte de la varianza total. Se empleó un gráfico de barras para representar la participación de las enfermedades en cada una de las dimensiones, y se creó una nueva base con las variables reducidas que será empleada en la siguiente etapa.

La base reducida generada en R, producto del ACP, es cargada en Colab para construir un predictor basado en redes neuronales del tipo Long Short-Term Memory. Para la construcción de la red neuronal, se inició con el ordenamiento de la base de datos por país y luego por año, seguido de la generación de las denominadas ventanas temporales, que fueron fijadas a 10 años, en donde cada país se toma de manera independiente, ordenado de forma cronológica. La función objetivo (target) de cada ventana es el año inmediatamente siguiente; por ejemplo, si se toma la ventana de 1990 a 1999, predice el 2000; se ajusta y repite el análisis de 1991 a 2000, y así sucesivamente hasta llegar a 2022, que es el año objetivo.

Con la conversión de la base a tipo arrays NumPy y la codificación de cada país con la técnica de one-hot encoding, se divide la base en un 80% para entrenamiento y un 20% para validación. La arquitectura de la red neuronal está formada por 100 etapas de entrenamiento con un tamaño de lote de 16 ejemplos por iteración, utilizando la función

fit() para ajustar los parámetros del modelo, que serán medidos mediante la métrica del Error Cuadrático Medio (MSE). Con el modelo de mayor rendimiento definido, se construyó una nueva base de datos con las predicciones generadas por la red neuronal y se realizó un ACP inverso en R utilizando las matrices del ACP original, para finalmente aplicarle una desestandarización.

Para el segundo predictor, a las variables name y continente se les aplicó un label encoding, y se procedió a dividir la base en la misma proporción que la anterior para su validación. El modelo de Random Forest, empleó el algoritmo ranger que lo entrena con una arquitectura de 1000 árboles, un número de variables de 4 (mtry), que determina que cada vez que el modelo debe hacer una división en un árbol selecciona 4 variables de forma aleatoria, y un tamaño mínimo de nodo de 1. Estos hiperparámetros fueron generados a partir de un código de búsqueda con la librería caret::train(). Establecido el modelo, se calcularon las métricas de Error Cuadrático Medio (MSE), Raíz del Error Cuadrático Medio (RMSE), Error Absoluto Medio (MAE) y Coeficiente de Determinación ( $R^2$ ). Finalmente, se empleó este modelo para crear una base de datos con las predicciones de cada una de las 32 enfermedades para los 199 países, dichas observaciones fueron las empleadas para realizar visualizaciones como series de tiempo e identificar las tendencias y las causas más influyentes en la mortalidad.

## RESULTADOS

El proceso de depuración de la base de datos consistió en una serie de procedimientos que inició con la verificación y descarte de registros duplicados mediante la función get\_dupes(), en lo que respecta a las variables categóricas, se retiraron símbolos que dificultan el procesamiento de datos, tales como la tilde y la virgulilla de la letra ñ, también se estableció que tanto países y continentes inicien con la primera letra mayúscula, a esta etapa se la conoce como coerción de datos categóricos. La base original no incorporaba información de la población ni del continente, por lo que, al incluirse mediante la unión con otras bases de datos, de una población de 6528 datos se generaron 152 faltantes que representan el 2.3% de los datos totales, que fueron eliminados para tener una base completamente limpia.

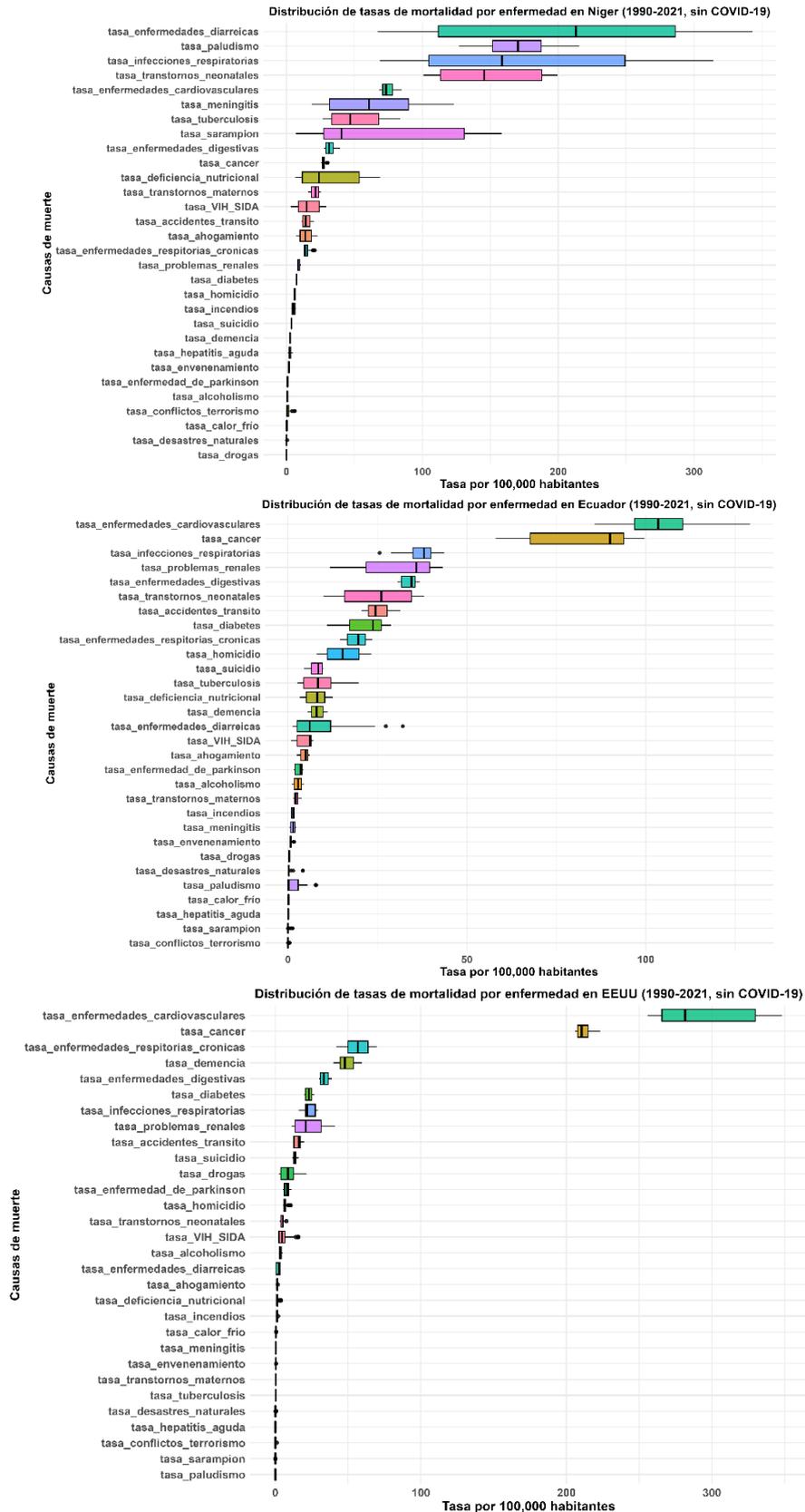


Figura 1. Diagramas de caja de tasas de mortalidad por 100.000 habitantes para las principales causas de muerte en tres países representativos, 1990–2021.

La visualización de datos partió con diagramas boxplot para todos los países con todas sus razones de muerte, de las cuales se han seleccionado a tres países con diferentes niveles de desarrollo: Níger que representa a los países de extrema pobreza, Ecuador como país en vías de desarrollo y Estados Unidos como país desarrollado. La elección responde a una estrategia comparativa basada en la diferencia de ingresos que permite contrastar los perfiles de mortalidad asociados a condiciones sociales, acceso a sistemas de salud, e infraestructura sanitaria.

La Figura 1 muestra los diagramas de cajas y bigotes por cada 100.000 habitantes de las principales causas de muerte en el período 1990–2021 para cada uno de los tres países seleccionados. La posición vertical ubica a las enfermedades de acuerdo con su impacto promedio, es decir, en la parte superior se encuentran aquellas con mayor peso relativo respecto al total de defunciones.

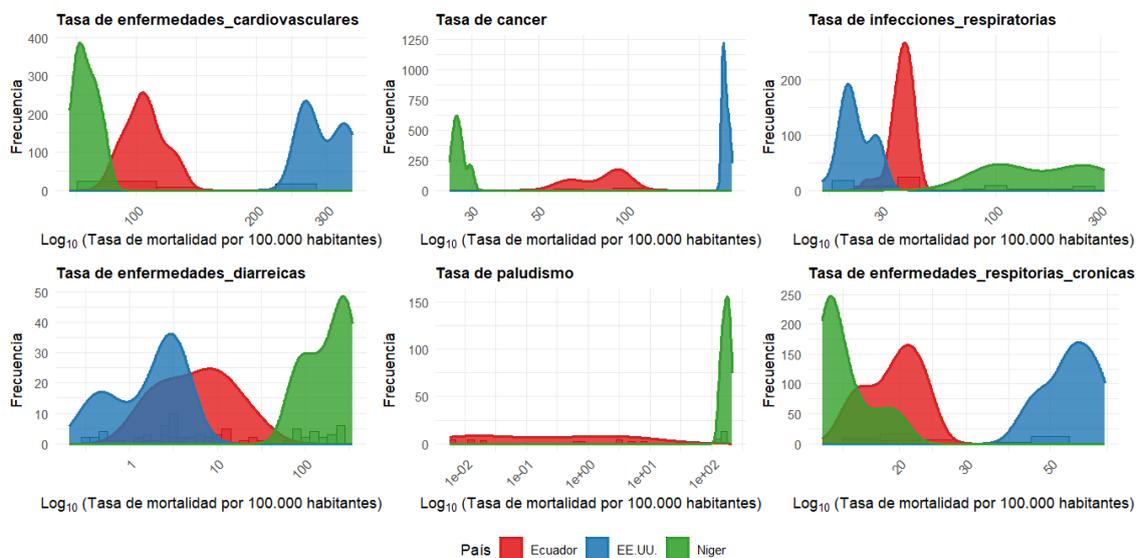
En Níger, se observa una supremacía de las enfermedades diarreicas, con un alto rango intercuartílico que evidencia una considerable variabilidad en los datos. Esto sugiere que pudo haber existido la presencia de brotes epidémicos, colapsos sanitarios, o la ausencia de campañas preventivas sostenidas, lo que genera repuntes frecuentes de esta enfermedad. Las otras tres causas de muerte que presentan una mediana cercana a las 300 muertes por cada 100 000 habitantes son el paludismo, las infecciones respiratorias y los trastornos neonatales.

En Ecuador, a primera vista, se diferencia medianas más bajas que en el país anterior, donde las enfermedades cardiovasculares constituyen la principal causa de muerte, con una baja variabilidad. El cáncer y las infecciones respiratorias completan la lista de las tres principales causas. La elevada cantidad de muertes ocasionadas por accidentes de tránsito (25 por cada 100 000 habitantes) evidencia deficiencias en la educación vial o carencias en la infraestructura del sistema vial. En general, la dispersión de las tasas es moderada, lo que refleja un sistema sanitario parcialmente consolidado.

Para Estados Unidos, la mortalidad se concentra en dos enfermedades: afecciones cardiovasculares y cáncer, con medianas superiores a las 200 muertes por cada 100 000 habitantes. En general, los rangos intercuartílicos de todas las enfermedades son bajos, lo que refleja un excelente desempeño de los sistemas de salud. A diferencia de los dos países anteriores, se observa una mayor presencia de enfermedades relacionadas con trastornos neurológicos y mentales.

Para avanzar con el análisis comparativo de los perfiles de mortalidad, se construyeron histogramas transformados a escala logarítmica para las seis enfermedades más representativas de los tres países. La selección se realizó tomando las tres causas con mayor mediana en cada país; al existir enfermedades coincidentes, se obtuvieron seis patologías en total: enfermedades cardiovasculares, cáncer, infecciones respiratorias, enfermedades diarreicas, paludismo y enfermedades respiratorias crónicas. En la Figura 2 se muestran las distribuciones de tasas de mortalidad ( $\log_{10}$  de muertes por cada 100 000 habitantes) para

estos tres países en el periodo 1990-2021, además, se incluye la correspondiente curva de densidad ajustada.



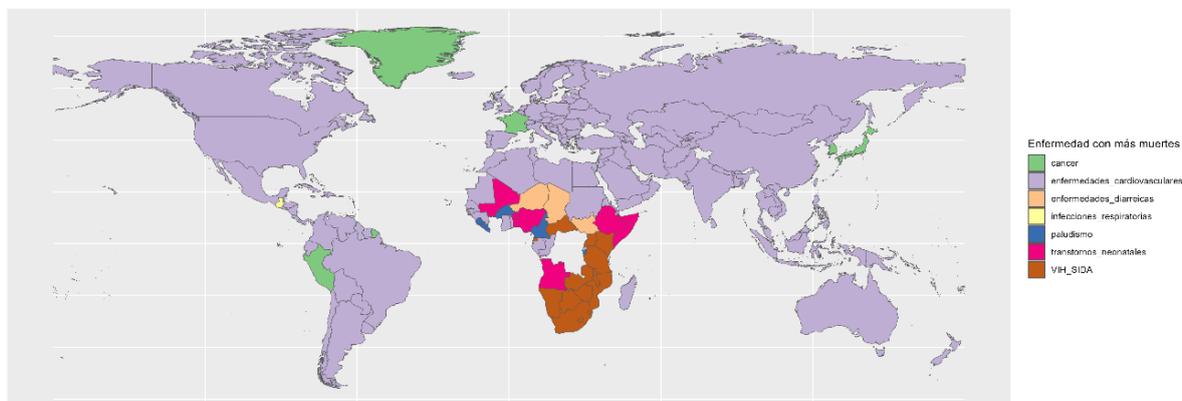
**Figura 2. Histogramas comparativos con curvas de densidad de tasas de mortalidad ( $\log_{10}$  por 100.000 habitantes) para seis causas seleccionadas, en Ecuador, EE.UU. y Níger (1990-2021).**

Las enfermedades cardiovasculares presentan distribuciones diferenciadas, para EE.UU. se muestra una distribución bimodal que representaría la existencia de dos periodos con altos brotes, sus picos se encuentran en torno a tasas de valores elevados (250-300), lo que podría indicar la coexistencia de dos períodos distintos o de dos subpoblaciones diferenciadas dentro del país. Ecuador y Níger presentan distribuciones con menor dispersión con valores mucho menores que los presentados por Estados Unidos.

En el caso del cáncer y de las enfermedades respiratorias crónicas se mantienen la tendencia, en el cual, Estados Unidos presenta la mayor cantidad de muertes, seguida de Ecuador y de Níger, pero es importante diferenciar que, en el caso del cáncer, a pesar de que es la segunda causa de muerte para Estados Unidos y Ecuador, sus frecuencias son completamente distintas, a tal punto que, la de Estados Unidos prácticamente quintuplica a la de Ecuador. En cuanto a las infecciones respiratorias, Níger y Ecuador se distribuyen en valores moderados a altos, con una notable presencia de Níger en el extremo superior.

EE.UU., en cambio, presenta una posible bimodalidad, con acumulaciones en dos rangos distintos, lo que podría reflejar cambios en las estrategias de control o en la evolución del comportamiento respiratorio a lo largo del período. Las enfermedades diarreicas exhiben una marcada diferencia: Níger se ubica claramente en tasas elevadas, mientras que Ecuador y EE.UU. se concentran en valores significativamente menores, con leves diferencias entre ellos. Esto muestra una posible vulnerabilidad en el acceso a agua potable y saneamiento en dicho país. En el caso del paludismo, la curva de Níger domina completamente el extremo superior, mientras que Ecuador y EE.UU. se agrupan en el rango más bajo posible. Este contraste evidencia la naturaleza endémica del paludismo en regiones africanas y su correcto control en otras ubicaciones geográficas.

Para tener una visualización más global de las tasas anuales y distribuciones por país, se estimó la carga acumulada de mortalidad por cada una de las 32 causas de muerte, para cual, se sumó el total de fallecimientos por país durante el período 1990-2021. La Figura 3 se presenta un mapa categórico en el que se identifica la causa con mayor número de muertes para cada país.



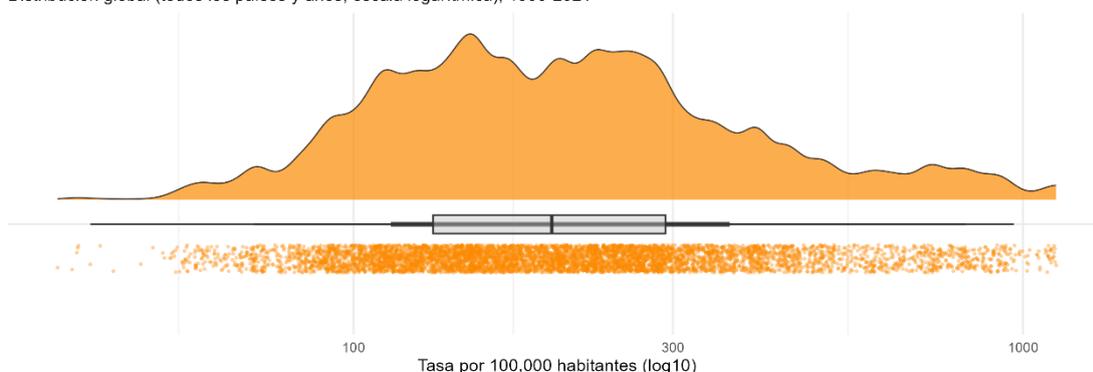
**Figura 3. Causa de muerte con mayor número acumulado de fallecimientos por país (1990-2021).**

Los resultados revelan patrones geográficos coherentes con las condiciones epidemiológicas globales: en países con buenos sistemas sanitarios como Estados Unidos, Europa, la mayoría de Asia y América del Sur, las enfermedades cardiovasculares se establecen como la principal causa de muerte, mientras que, en gran parte de África hay un predominio de causas ligadas a enfermedades infecciosas como el paludismo, enfermedades diarreicas, VIH/SIDA y los trastornos neonatales. Con base a estos resultados, se seleccionaron las tres enfermedades con mayor participación que son: enfermedades cardiovasculares, VIH/SIDA y paludismo.

La construcción de los denominados gráficos raincloud (Figura 4) complementan el análisis espacial de las tres mayores causas de muerte acumulada, estos gráficos combinan una estimación suavizada de la densidad (curva superior), un diagrama de cajas (centro) y una dispersión puntual (inferior), representando simultáneamente forma, tendencia y variabilidad de las tasas de mortalidad transformada a escala logarítmica ( $\log_{10}$ ).

#### Raincloud plot - Mortalidad por Enfermedades cardiovasculares

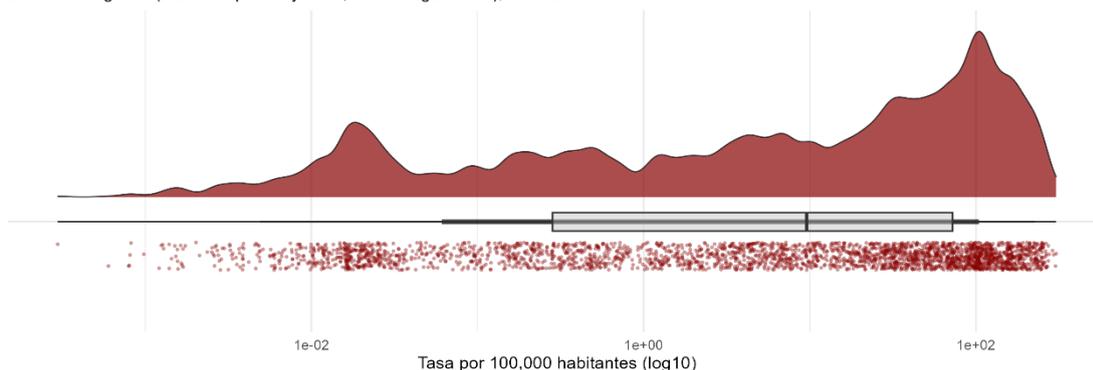
Distribución global (todos los países y años, escala logarítmica), 1990-2021



Fuente: Elaboración propia con base en GBD (2021)

**Raincloud plot - Mortalidad por Paludismo**

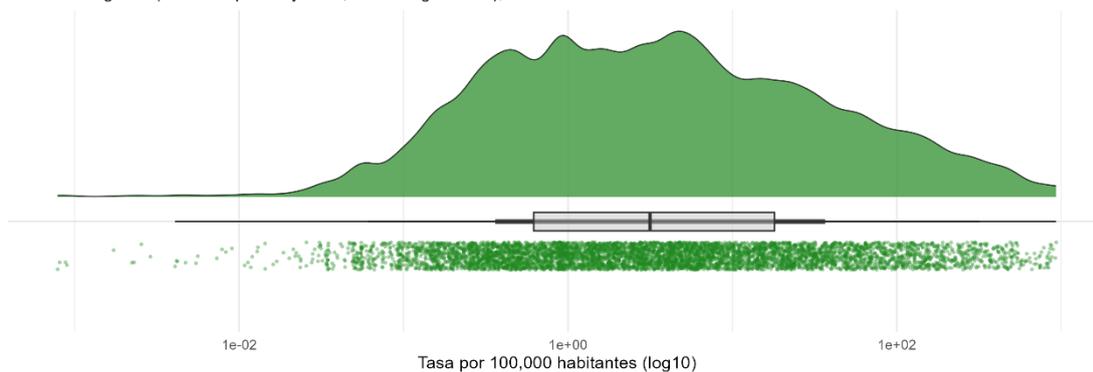
Distribución global (todos los países y años, escala logarítmica), 1990-2021



Fuente: Elaboración propia con base en GBD (2021)

**Raincloud plot - Mortalidad por VIH/SIDA**

Distribución global (todos los países y años, escala logarítmica), 1990-2021



Fuente: Elaboración propia con base en GBD (2021)

**Figura 4. Gráficos raincloud de tasas de mortalidad por cada 100.000 habitantes (escala  $\log_{10}$ ) para (a) enfermedades cardiovasculares, (b) VIH/SIDA y (c) paludismo, en todos los países y años (1990–2021).**

La Figura 4a, corresponde al comportamiento de las enfermedades cardiovasculares, que muestra una distribución relativamente simétrica y unimodal, con valores concentrados en el intervalo  $\log_{10}(100)$  a  $\log_{10}(300)$  muertes por cada 100 000 habitantes, esta simetría sugiere que la mayoría de los países poseen tasas de mortalidad altas que oscilan cerca del promedio global. Su mediana se ubica ligeramente por encima de 200 con pocos valores atípicos. La Figura 4b, representa al VIH/SIDA que cuenta con una asimetría con sesgo positivo, con una mediana en  $1e+01$  (10) acompañada de una mayor presencia de valores atípicos, esta configuración sugiere que solo un subconjunto de países (principalmente africanos) concentran su mortalidad en esta enfermedad.

El paludismo representado en la Figura 4c, muestra una estructura más extrema y multimodal con una mediana muy baja, lo que indica que la mayoría de los países tienen mortalidad casi nula por esta causa, pero pocos registran tasas sumamente altas. Estos tres gráficos ilustran una diversidad estructural en la distribución de la mortalidad da nivel global, mientras las enfermedades cardiovasculares presentan una distribución homogénea y simétrica, el VIH/SIDA y el paludismo muestran distribuciones sesgadas mostrando una inequidad en la carga epidemiológica.

Como fase previa a aplicar las técnicas multivariantes, se procedió a analizar la estructura interna del conjunto de datos mediante la matriz de correlación de Pearson. Esta se construyó sobre los valores estandarizados ( $Z_{score}$ ) de las 32 causas de muertes en todos

los países y años disponibles. La Figura 5 muestra dicha matriz con los coeficientes codificados por intensidad de color, junto a sus valores numéricos correspondientes.

**Matriz de Correlación - Enfermedades (Z-score)**

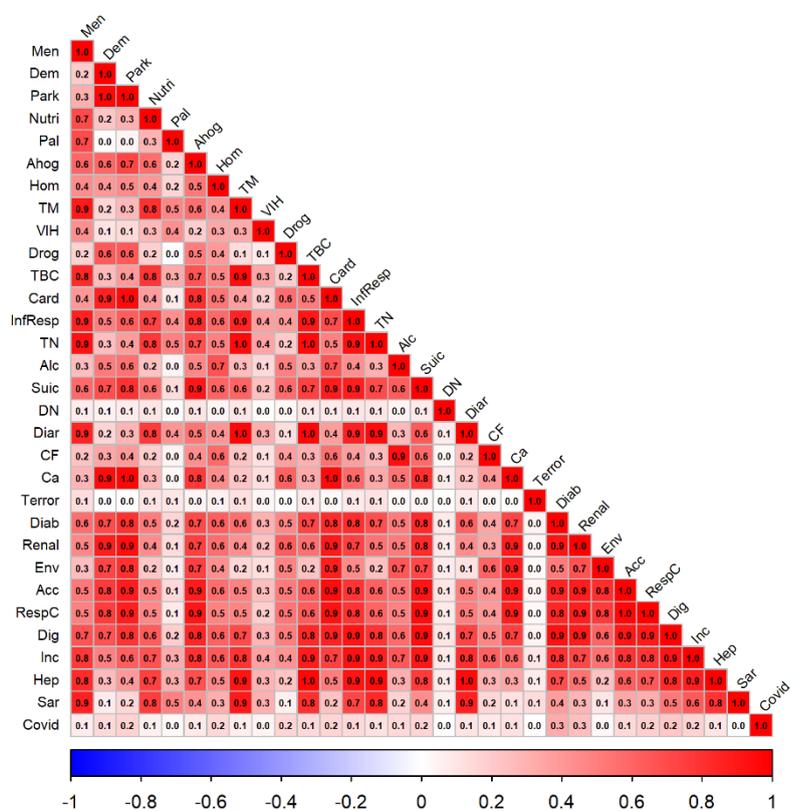


Tabla de Abreviaciones de Enfermedades

Enfermedad	Abreviación
meningitis	Men
demenia	Dem
enfermedad_de_parkinson	Park
deficiencia_nutricional	Nutri
paludismo	Pal
ahogamiento	Ahog
homicidio	Hom
trastornos_maternos	TM
VIH_SIDA	VIH
drogas	Drog
tuberculosis	TBC
enfermedades_cardiovasculares	Card
infecciones_respiratorias	InfResp
trastornos_neonatales	TN
alcoholismo	Alc
suicidio	Suic
desastres_naturales	DN
enfermedades_diarreicas	Diar
calor_frio	CF
cancer	Ca
conflictos_terrorismo	Terror
diabetes	Diab
problemas_renales	Renal
envenenamiento	Env
accidentes_transito	Acc
enfermedades_respiratorias_cronicas	RespC
enfermedades_digestivas	Dig
incendios	Inc
hepatitis_aguda	Hep
sarampion	Sar
covid_19	Covid

**Figura 5. Matriz de correlación de Pearson entre las principales causas de muerte (Z-score), calculada sobre todos los países y años del período 1990–2021.**

Se observan agrupaciones evidentes entre enfermedades no transmisibles, por ejemplo, las enfermedades cardiovasculares presentan correlaciones elevadas con la diabetes (0.8), con enfermedades renales (0.9), y con enfermedades digestivas (0.9). Asimismo, se identifican correlaciones relevantes entre las enfermedades respiratorias crónicas y cardiovasculares, y relaciones muy bajas con enfermedades como el VIH/SIDA (0.2), paludismos (0.1) y enfermedades diarreicas (0.1).

La presencia de correlaciones altas y consistentes entre múltiples causas de muerte evidencia una redundancia estadística que requiere la aplicación de una técnica que permitirá reducir la dimensionalidad del conjunto de datos. El análisis de componentes principales (ACP) es ideal para hacerlo, ya que, a más de reducir, identifica comportamientos epidemiológicos sin pérdida significativa de información.

Previo a la aplicación del Análisis de Componentes Principales (ACP), se hace necesario la comprobación del cumplimiento de sus supuestos, para lo cual se inicia con la realización de una prueba de bondad de ajuste mediante los gráficos Q–Q a los valores estandarizados por Z-score. La Figura 6 muestra los gráficos correspondientes a las 31 causas de muerte analizadas (excluyendo COVID-19), comparando los cuantiles teóricos de una distribución



normal con los cuantiles muestrales para cada variable.

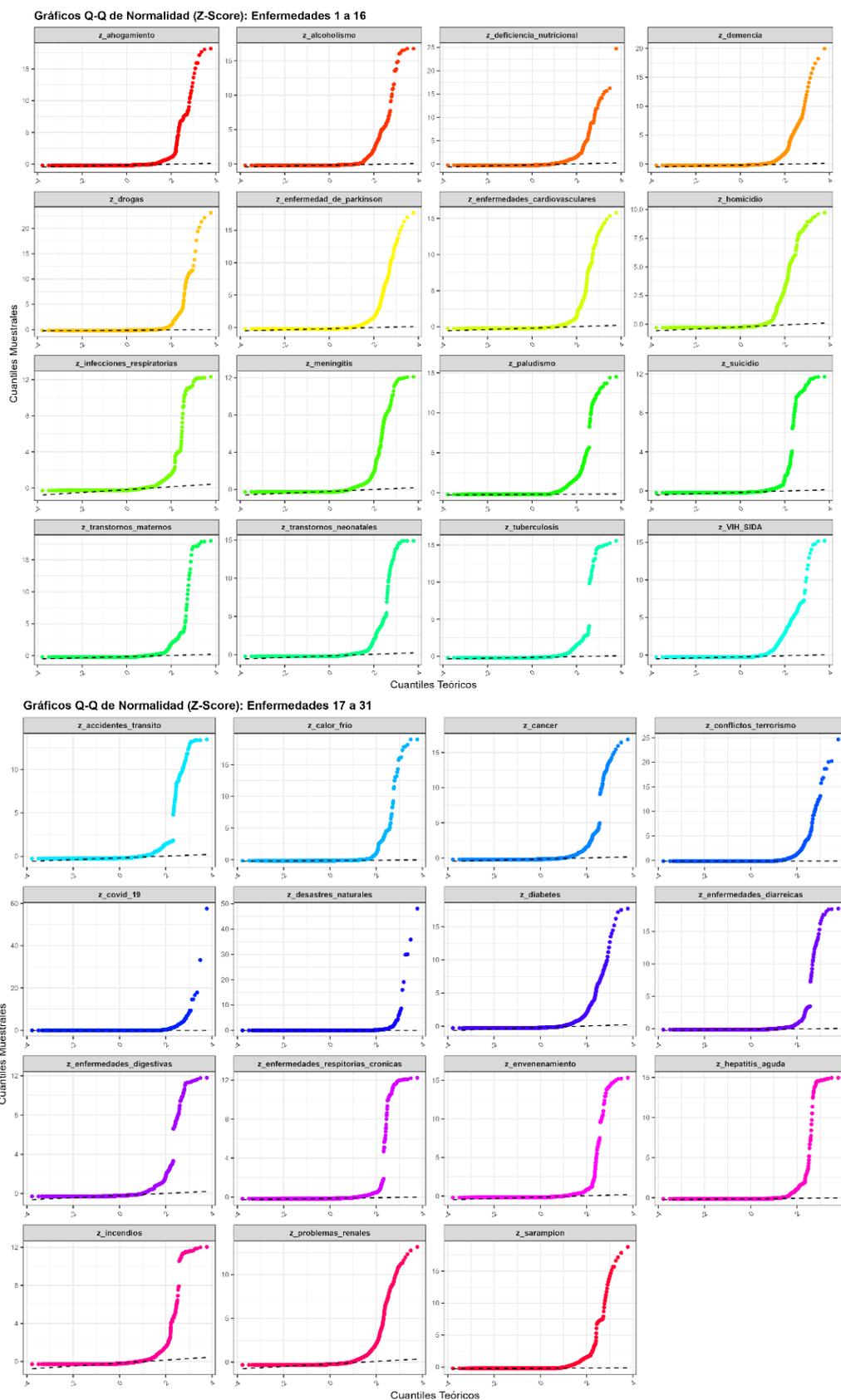


Figura 6. Gráficos Q-Q para la evaluación de la normalidad marginal de cada variable estandarizada (Z-score).

En general, se observa que los cuantiles centrales se alinean razonablemente con la diagonal teórica en la mayoría de las enfermedades, lo que sugiere una aproximación aceptable a la normalidad. No obstante, también se evidencian colas que se alejan en varias distribuciones (especialmente en enfermedades como VIH/SIDA, sarampión, o desastres naturales) lo que indica cierta asimetría o presencia de valores extremos. A pesar de estas desviaciones, el ACP clásico es reconocido como una técnica robusta ante variaciones moderadas del supuesto de normalidad multivariada, por lo que, se pueda dar como válido este supuesto.

Al trabajar sobre los datos estandarizados ya se garantiza que la media de las variables sea cero (segundo supuesto) y una dispersión simétrica en torno a este valor, asegurando, además, que ninguna variable domine la varianza total por su escala o magnitud, y permita comparar el peso relativo de cada componente de forma equitativa.

El tercer supuesto estadístico para la aplicación del ACP es la verificación de varianzas iguales, para lo cual se realizó la prueba de Brown-Forsythe, cuyo objetivo es contrastar la homogeneidad de varianzas entre grupos, esta prueba es una alternativa robusta a la prueba de Levene, especialmente cuando las distribuciones presentan asimetrías ya que trabaja con la mediana (20). La hipótesis nula ( $H_0$ ) de esta prueba plantea que las varianzas son iguales entre grupos para cada variable evaluada, en el presente estudio, todas las casas de muerte presentaron valores p menores a 0.05, lo cual lleva al rechazo de la hipótesis nula y concluir que existen diferencias significativas en la dispersión de las variables, esto refuerza la decisión metodológica de utilizar la matriz de correlaciones estandarizada como base del ACP, ya que, la matriz de covarianzas es inapropiada ante la presencia de varianzas heterogéneas.

Finalmente, se evaluó si la muestra es representativa y apropiada (adecuación muestral) mediante el índice de Kaiser-Meyer-Olkin (KMO), el cual mide la proporción de varianza común entre las variables, en términos generales el índice fue alto (mayor a 0.8 en promedio) lo que implica que el conjunto de datos es altamente apropiado para análisis estadísticos posteriores. La Figura 7 muestra los valores individuales del KMO por variable, destacando que las variables como deficiencia nutricional, enfermedades digestivas o respiratorias crónicas superan el 0.90, sin embargo, el paludismo registró el valor más bajo (0.62) que podría deberse a una distribución geográfica muy focalizada de esta enfermedad, pero que a pesar de ser la de menor valor, se encuentra dentro de rangos aceptables.

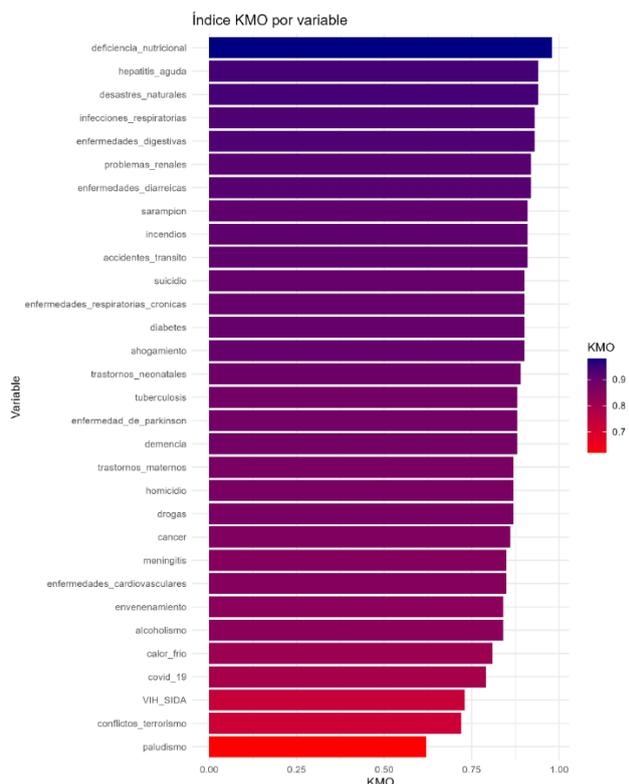


Figura 7. Índice KMO por variable.

Luego de la comprobación de los supuestos, se aplicó un Análisis de Componentes Principales sobre las variables estandarizadas (Z\_score), la Figura 8, se muestra la varianza explicada en cada una de las dimensiones resultantes.

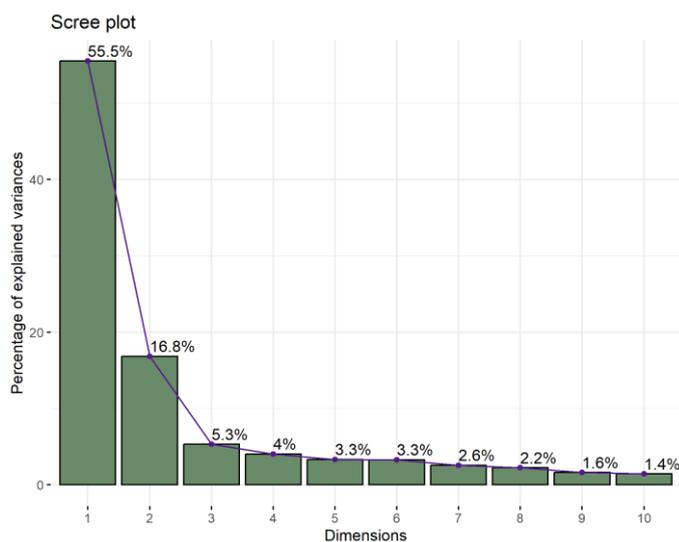
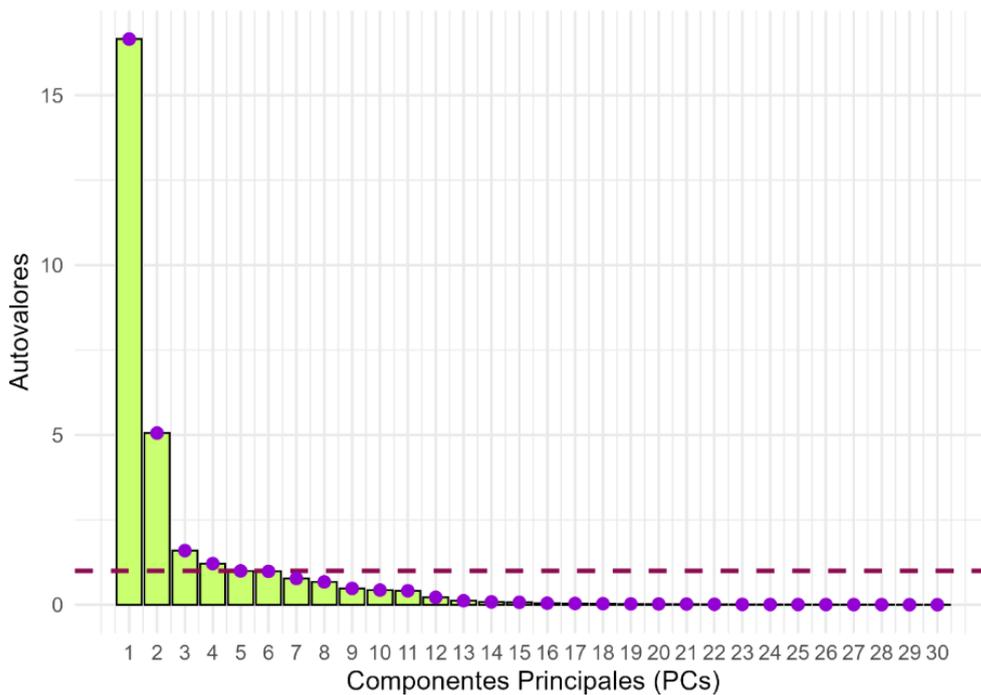


Figura 8. Scree plot del Análisis de Componentes principales (ACP): porcentaje de varianza explicada por dimensión.

Para la selección del número de componentes principales se sustentó en el criterio de Kaiser que se basa en conservar las componentes con valores propios mayores a 1, basado en el análisis gráfico de la Figura 9, se decide mantener 5 componentes principales que explican aproximadamente el 84.9% de la varianza acumulada.

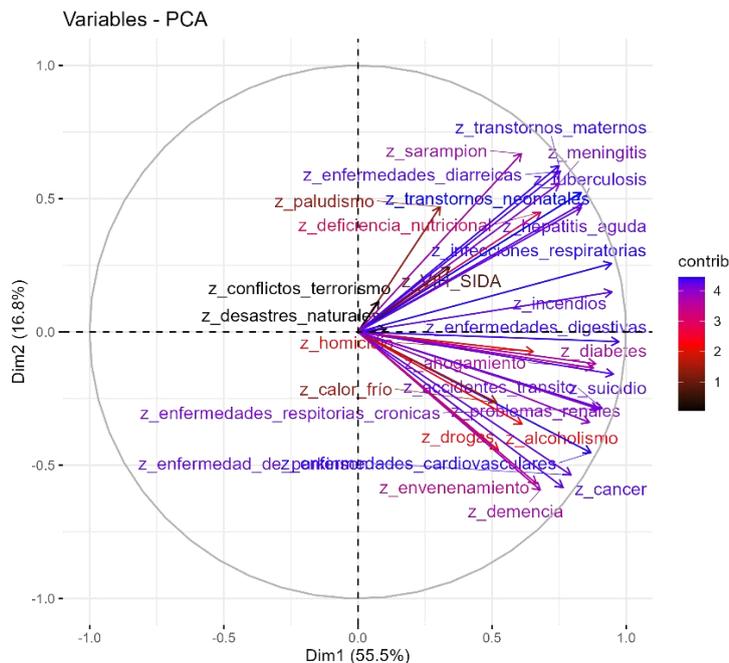


### Criterio de Kaiser - Selección de Componentes Principales



**Figura 9. Selección de componentes principales según el criterio de Kaiser.**

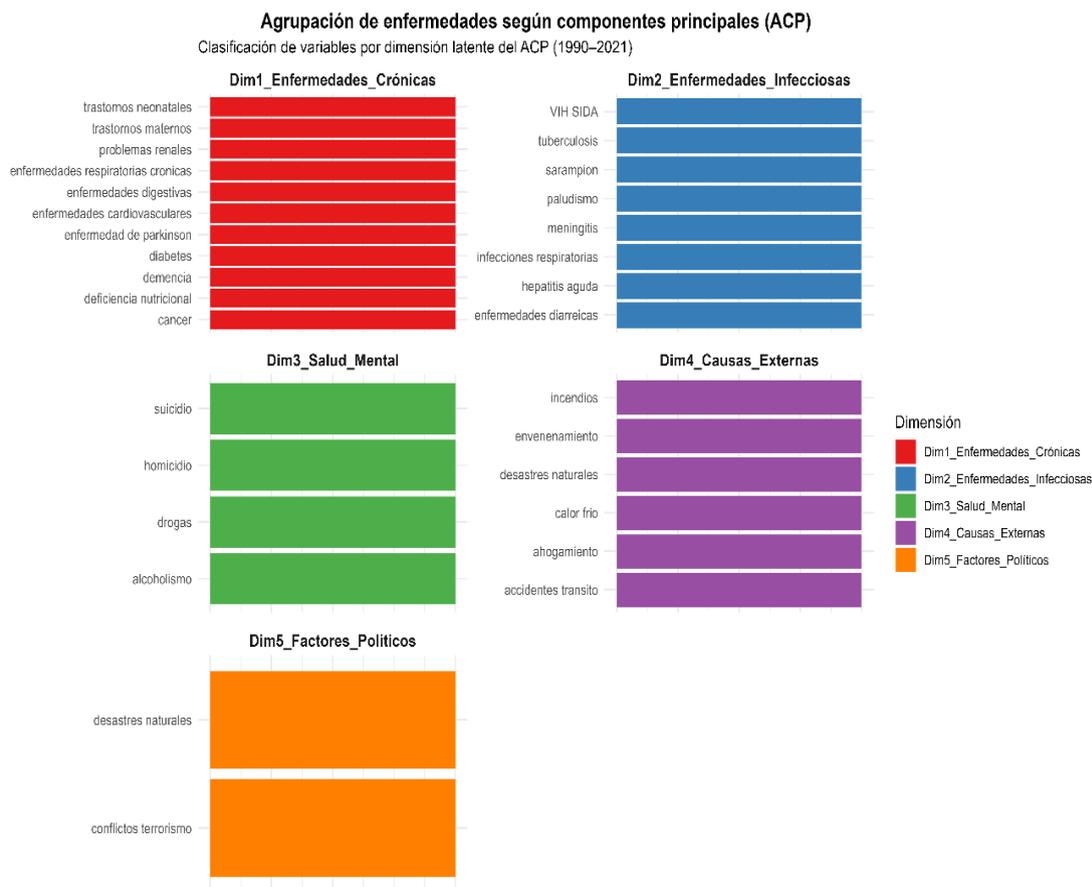
En la Figura 10 se presenta el círculo de correlaciones correspondiente al ACP, el cual representa la relación entre las variables estandarizadas (enfermedades) y los dos primeros componentes retenidos.



**Figura 10. Círculo de correlaciones del ACP: contribución de las enfermedades en las dos primeras dimensiones (72.3% de varianza explicada).**

El eje horizontal (Dim1) explica el 55.5% de la varianza total, mientras que el eje vertical (Dim2) un 16.8%, en donde cada flecha representa una enfermedad en específico, la





**Figura 12. Agrupación de enfermedades por dimensión latente según el Análisis de Componentes Principales (1990–2021).**

La agrupación final para la reducción de dimensiones queda establecida por la Figura 12, donde la primera dimensión acapara la mayor cantidad de enfermedades, y, además, contiene a las dos principales causas de muerte que dejó al descubierto el estudio exploratorio.

Una vez definida la base de datos reducida, se procedió al diseño de un modelo de predicción basada en redes neuronales del tipo Long Short-Term Memory (LSTM), en la cual se estableció una secuencia de ventanas temporales cada 10 años para cada país, con lo cual cada ventana predice el siguiente año. El modelo fue implementado en TensorFlow/Keras y entrenado con una arquitectura de tres capas LSTM con 200, 100 y 50 neuronas cada una, además, cada capa consta con un dropout para evitar el sobreajuste con una función de activación tipo ReLU. La capa de salida está provista de cinco neuronas, una por cada dimensión de la base de datos (Figura 13).

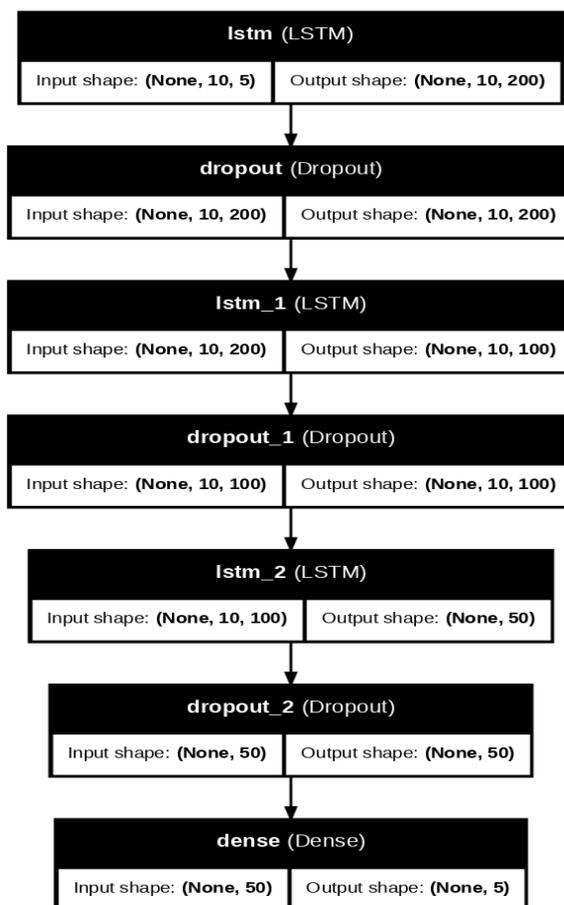


Figura 13. Arquitectura de red neuronal LSTM.

Definida la arquitectura, se dividió a la base en dos partes, el 80% para entrenamiento y 20% para validación, el modelo fue compilado con optimizador Adam y la función de pérdida fue la del error cuadrático medio (MSE). Finalmente, el entrenamiento constó de 100 épocas con lotes de 16 observaciones.

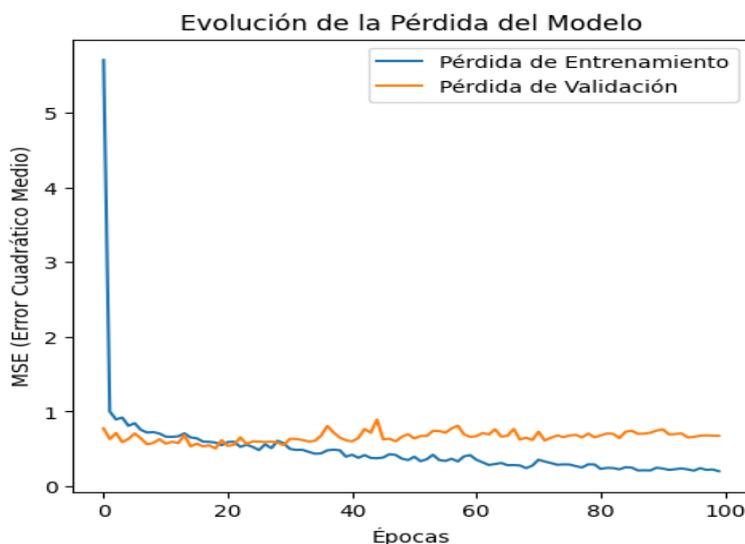


Figura 14. Evolución de la pérdida del modelo durante el entrenamiento.

La Figura 14 muestra la evolución de la función de pérdida del modelo neuronal durante el proceso de entrenamiento, utilizando el error cuadrático medio (MSE) como métrica de desempeño. Durante las 10 primeras épocas de entrenamiento se observa una rápida disminución, seguida de una fase de estabilidad con pequeñas oscilaciones, y dado que, la diferencia entre ambas curvas es relativamente igual a partir de las 60 épocas de entrenamiento se concluye que el modelo no presenta sobreajustes severos.

Para medir el desempeño del modelo neuronal, se compara las respuestas del sistema frente a los datos de evaluación, con lo cual, alcanzó un MSE final de 0.6740, un error absoluto medio (MAE) de 0.2108 y un coeficiente de determinación  $R^2$  de 0.5246, basado en estas métricas se concluye que el modelo presenta un desempeño predictivo moderado, con capacidad para explicar aproximadamente el 52% de la variabilidad para el año 2022.

Aunque el valor de  $R^2$  no sugiere una predicción perfecta, resulta estadísticamente aceptable considerando la alta complejidad, variabilidad y multicausalidad de los patrones de mortalidad global. El bajo MAE refleja una buena precisión promedio en términos absolutos, lo cual es favorable dado el carácter multivariado del modelo.

Una vez obtenidas las predicciones de las cinco dimensiones principales mediante la red neuronal LSTM, se reconstruyó las tasas individuales de mortalidad a través de un ACP inverso, es decir, se pasó de las cinco dimensiones predichas a las 31 causas de muertes originales que corresponden al año 2022, para finalmente aplicar una desestandarización y obtener observaciones en escalas reales. Aunque las métricas de desempeño en el conjunto de prueba resultaron aceptables (MAE = 0.21;  $R^2$  = 0.52), la aplicación directa de este procedimiento generó algunos valores negativos en enfermedades específicas, lo cual carece de sentido.

Este comportamiento puede explicarse por la naturaleza lineal del ACP y lejos de invalidar el proceso, este resultado ilustra las limitaciones inherentes a técnicas lineales cuando se enfrentan a relaciones complejas y distribuciones asimétricas, por tal razón, se optó por no representar ni analizar estos valores reconstruidos.

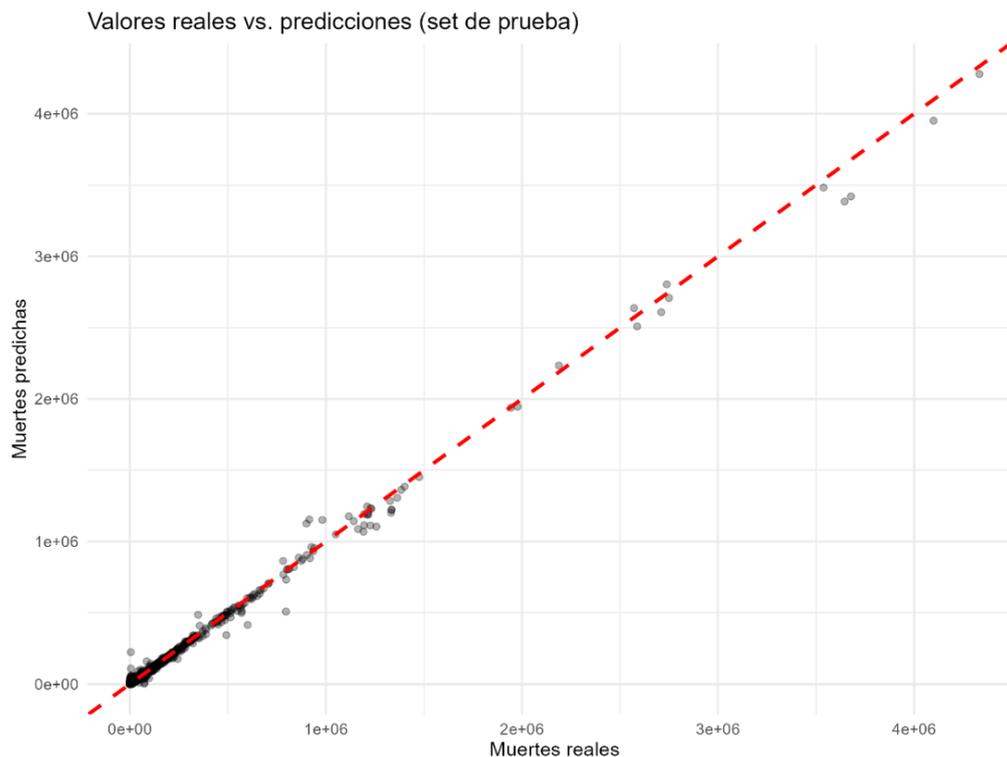
Como alternativa a la red neuronal y para evitar los inconvenientes anteriores, se implementó un modelo de Random Forest directamente sobre la estructura original de datos, sin necesidad de reducir previamente la dimensionalidad. Esta aproximación permite modelar relaciones no lineales y manejar la real interacción entre variables.

A la base original que contiene la información de las 32 causas de muerte se le aplicó un cambio a formato largo, con lo cual, cada fila representa una combinación única de país, año, causa de muerte, población y continente, además, es importante recordar que se elimina al COVID-19 para evitar sesgos por comportamientos atípicos que puedan afectar el desempeño del modelo predictor. Finalmente, para definir a todas las variables predictoras, se procedió a codificar por etiquetas (label encoding) a cada uno de los países.

Tras definir los hiperparámetros óptimos con validación cruzada —específicamente  $mtry = 4$  y  $min.node.size = 1$ — se entrenó el modelo definitivo con 1 000 árboles utilizando el algoritmo ranger, conocido por su eficiencia y escalabilidad. La partición de los datos se

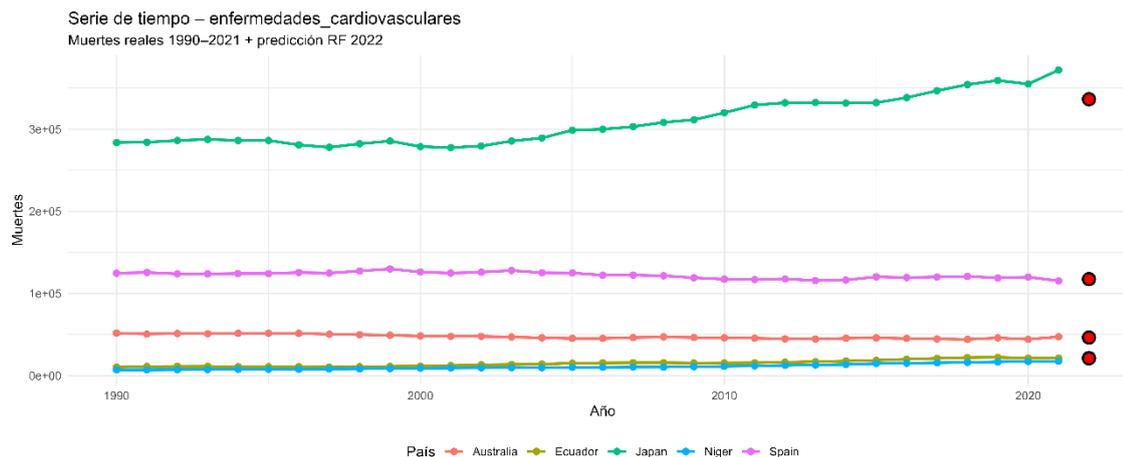
realizó asignando un 80% para entrenamiento y un 20% para prueba.

A diferencia de otras técnicas multivariantes que presentan problemas ante colinealidad y multicolinealidad entre variables, el algoritmo Random Forest es altamente robusto, ya que selecciona subconjuntos aleatorios de variables en cada división. Esto permite aprovechar la información redundante sin degradar el rendimiento predictivo, lo que resulta especialmente útil en contextos multicausales como el estudio de la mortalidad global, donde las variables sanitarias, demográficas y geográficas suelen estar fuertemente interrelacionadas.



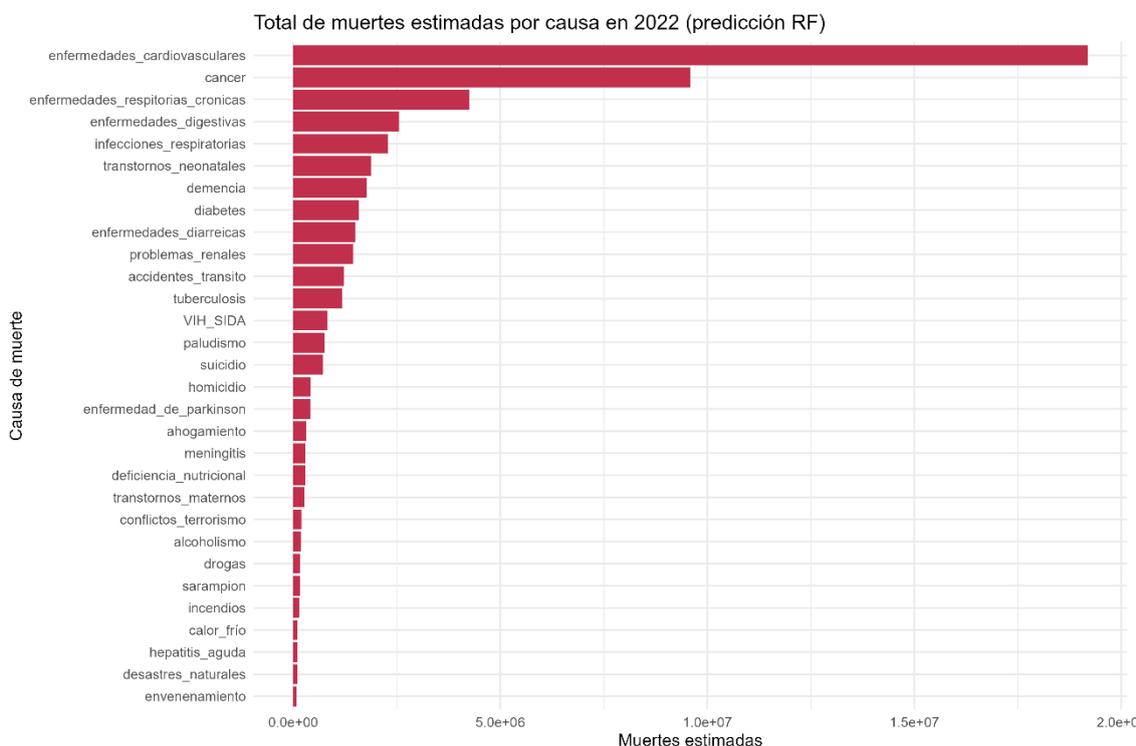
**Figura 15. Comparación entre valores reales y predicciones del modelo Random Forest en el conjunto de prueba.**

El rendimiento del modelo de Random Forest se visualiza en la Figura 15, y muestra un correcto desempeño reflejado por su coeficiente de determinación  $R^2 = 0.9958$  que indica que el modelo cuenta con una capacidad explicativa del 99.58% de la varianza total, por parte del error cuadrático medio (MSE) se obtuvo un valor de 22 327.536 y el error absoluto medio (MAE) alcanzó los 512.30 que representa la desviación media absoluta entre las predicciones y los valores reales, y a diferencia el MSE su penalización es menos severa a errores grandes, por lo que es una métrica más robusta.



**Figura 16. Serie temporal de muertes por enfermedades cardiovasculares (1990–2021) y predicción para 2022 mediante Random Forest en cinco países representativos.**

Terminado el entrenamiento y la validación del modelo de Random Forest, se procedió a emplear el modelo para predecir las muertes en el año 2022. Estas predicciones permiten extender el análisis temporal más allá de los datos históricos observados (1990–2021), y son particularmente útiles para monitorear tendencias recientes o emergentes. A modo de ejemplo, la Figura 16 muestra la serie temporal de muertes por enfermedades cardiovasculares en cinco países, uno por continente (Australia, Ecuador, Japón, Níger y España), incorporando la predicción correspondiente a 2022 representadas por los puntos rojos.



**Figura 17. Total global de muertes estimadas por causa – Predicción 2022 (RF).**

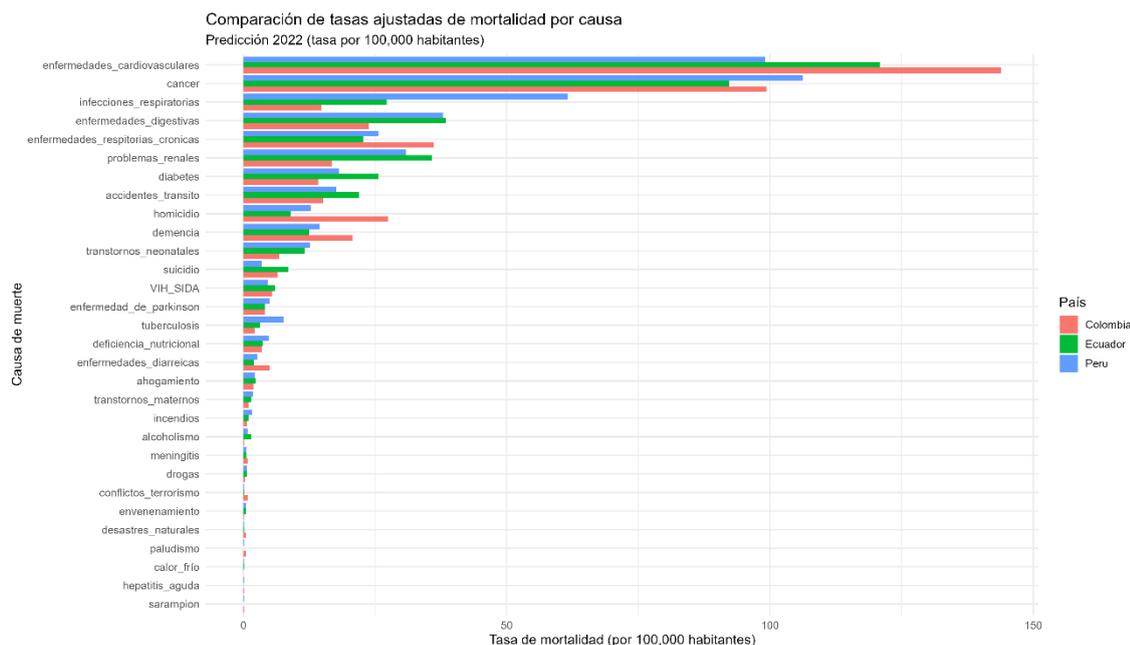
La cantidad de muertes estimadas para el año 2022, según el modelo Random Forest, se representa en la Figura 17. En ella se confirma que las enfermedades cardiovasculares encabezan la lista a nivel mundial, seguidas, a considerable distancia, por el cáncer y las enfermedades respiratorias crónicas. La única enfermedad asociada a la salud mental que aparece en los primeros puestos es la demencia, ubicado en la séptima posición. El listado de las diez principales causas de muerte se cierra con los problemas renales, seguidos muy de cerca por una causa externa: los accidentes de tránsito.

La Figura 18 presenta la evolución histórica de las muertes por homicidio en los cinco países con mayor carga acumulada entre 1990 a 2021, juntamente con sus predicciones para 2022 generadas por el modelo de bosque aleatorio, donde Colombia y Sudáfrica muestran tendencias decrecientes marcadas, mientras que Venezuela ha tenido un aumento considerable hasta alcanzar su pico máximo entre los años 2015 y 2016 que coincide con una de las mayores oleadas migratorias de dicho país. Países como El Salvador y Honduras exhiben picos abruptos y relativamente recientes, confirmando la existencia de patrones temporales de alta complejidad, por lo que, vale resaltar que las predicciones para 2022 se mantienen alineadas con la tendencia reforzando la estabilidad del modelo.



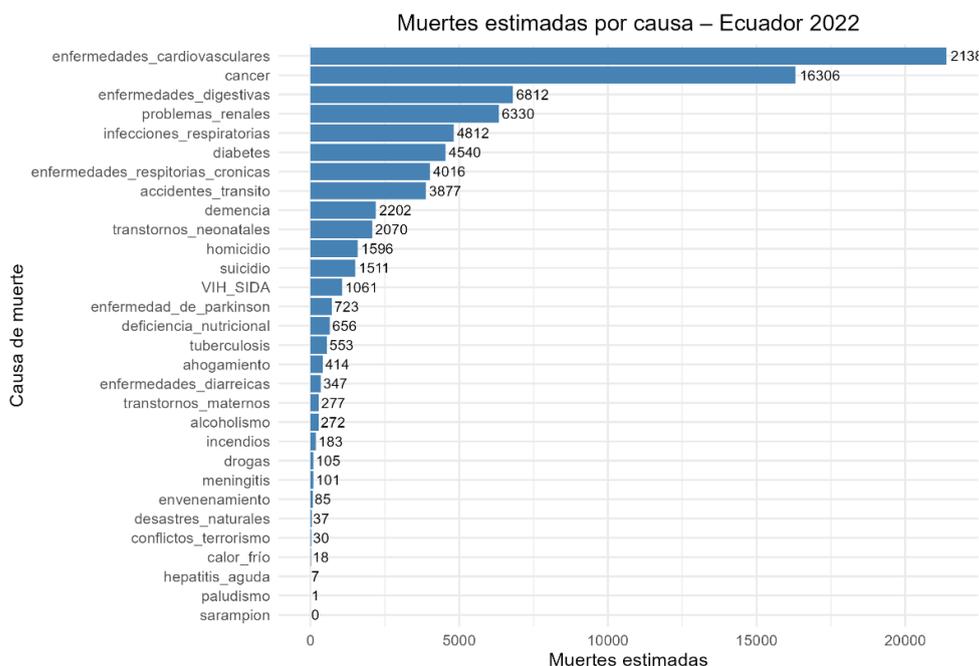
**Figura 18. Serie de tiempo de muertes por homicidio en los cinco países con mayor carga histórica (1990 - 2021) y predicciones RF para 2022.**

La comparación de las tasas ajustadas de mortalidad por cada cien mil habitantes se representa en la Figura 19, y corresponden a las predicciones del año 2022 para Colombia, Ecuador y Perú, países que fueron seleccionados por su cercanía geográfica, en la cual se observa que las enfermedades cardiovasculares se posicionan como la principal causa de muerte, resaltando a Colombia que supera los 140 casos por cada 100 mil habitantes, mientras que en el cáncer Perú presenta la tasa más elevada. Entre las diez primeras causas también se incluyen infecciones respiratorias, enfermedades de carácter digestivo, accidentes de tránsito y diabetes, lo cual indica que tanto las enfermedades crónicas como las causas externas representan una carga importante en la región.



**Figura 19. Comparación de tasas de mortalidad por causa entre Colombia, Ecuador y Perú para el año 2022 (predicción RF).**

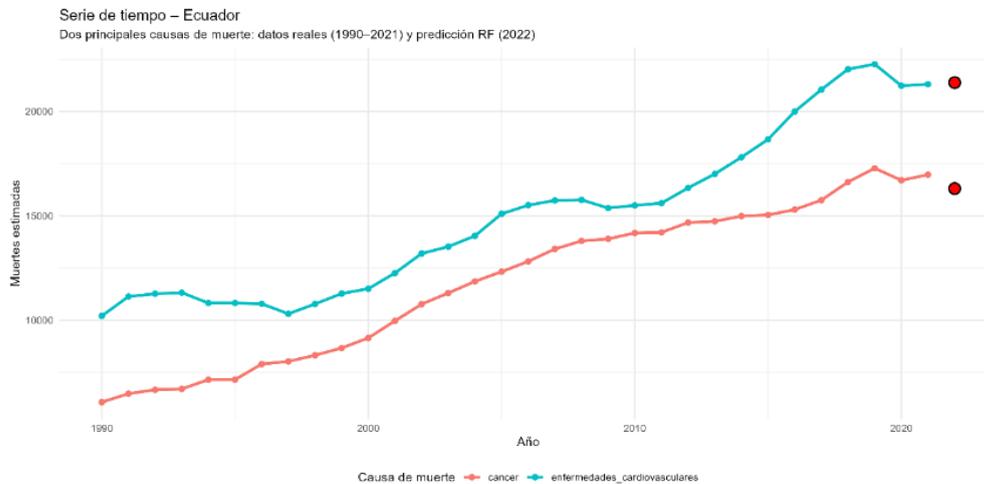
Para mostrar la estimación total de muertes en Ecuador para el año 2022, se emplea la Figura 20, en la cual se evidencia que las muertes cardiovasculares destacan como la principal causa de muerte con una cifra estimada cercana a los 21 000 decesos, seguida por el cáncer con 16 306 muertes, y las enfermedades digestivas con 6 812 decesos.



**Figura 20. Muertes estimadas por causa - Ecuador, predicciones 2022.**

La Figura 21 permite evidenciar de mejor manera el comportamiento histórico y las predicciones para 2022 de las dos principales causas de muerte para Ecuador, en la cual, hay una tendencia creciente a lo largo de los años, que podría explicarse en parte por el

incremento poblacional, el envejecimiento demográfico, y posibles factores asociados al estilo de vida, pero que también mostrarían la carencia o deficiencia de políticas públicas en combatir o mitigar las elevadas tasas de mortalidad por enfermedades cardiovasculares y cáncer.



**Figura 21. Comportamiento temporal de las dos principales causas de muerte en Ecuador.**

## DISCUSIÓN

Los resultados obtenidos en el presente confirman la utilidad de aplicar enfoques multivariantes para analizar patrones complejos de mortalidad a escala global. A diferencia de los análisis univariados tradicionales, la combinación de técnicas como el Análisis de Componentes Principales (ACP), redes neuronales LSTM (21) y modelos Random Forest permitieron captar la estructura latente de las causas de muerte, identificar agrupamientos epidemiológicos y generar predicciones con buenas métricas de rendimiento (22). Este enfoque integral no solo resalta diferencias geográficas, sino también diferencias estructurales entre países con distintos niveles de desarrollo social y económico.

Para evitar el sobreajuste en el modelo Random Forest, se implementaron técnicas de regularización específicas durante su configuración. Se utilizaron 1000 árboles para estabilizar las predicciones y se ajustó el parámetro *mtry* para limitar el número de variables consideradas en cada división, lo que favoreció la diversidad entre árboles. Además, se exploró el valor de *min.node.size* para controlar la profundidad de los nodos terminales y reducir la posibilidad de que el modelo se ajustara a ruido específico del entrenamiento. La validación se realizó sobre un conjunto de prueba independiente, empleando métricas ( $R^2 = 0.99$ ; MAE = 512), junto con visualizaciones de errores, lo que evidenció un buen desempeño general sin signos de sobreajuste.

Uno de los hallazgos centrales es la consolidación de las enfermedades cardiovasculares y el cáncer como principales causas de muerte global, especialmente en países de altos ingresos, lo cual coincide con estudios previos que evidencian la transición epidemiológica hacia enfermedades crónicas no transmisibles como factor dominante en la carga de enfermedad mundial (23). En contraste, en países de bajos ingresos como Níger, se observa la persistencia de enfermedades infecciosas como el paludismo o enfermedades diarreicas, lo que ratifica la influencia decisiva de las condiciones sanitarias, el acceso a servicios básicos y la eficacia

de los sistemas de salud pública (24).

El ACP permitió sintetizar la complejidad de 32 causas de muerte en cinco dimensiones epidemiológicas, ofreciendo un mapa conceptual del comportamiento conjunto de las enfermedades. Esta reducción de dimensionalidad no solo facilitó la interpretación de los datos, sino que constituyó la base para la implementación de modelos predictivos. La primera dimensión, que agrupa las enfermedades cardiovasculares, digestivas y renales, mostró una elevada varianza explicada, lo que sugiere un patrón común de enfermedades crónicas ligadas a estilos de vida. Este resultado es consistente con literatura que vincula estos padecimientos con factores de riesgo compartidos como el sedentarismo, la dieta y el envejecimiento poblacional (25).

El modelo Random Forest superó ampliamente en rendimiento a la red neuronal LSTM, tanto en capacidad explicativa ( $R^2 = 0.99$ ) como en precisión ( $MAE = 512$ ). Este resultado no solo valida la robustez del modelo ante relaciones no lineales, sino que también pone en evidencia las limitaciones de las redes neuronales cuando se aplican a fenómenos con alta heterogeneidad estructural y baja densidad de datos por grupo. Estos hallazgos refuerzan lo planteado por Fernández-Delgado en (26) quienes advierten que los modelos de árboles de decisión frecuentemente superan a redes neuronales en tareas tabulares complejas.

No obstante, es importante subrayar que incluso los modelos con buen desempeño pueden presentar fallas en contextos con eventos abruptos. Un ejemplo relevante es el caso de Ecuador, donde las proyecciones del modelo no anticiparon el fuerte incremento real de muertes por homicidio en años posteriores a 2021, ya que, hasta este año, no figuraba entre los países con más homicidios. Este desfase no implica un error del modelo, sino que pone en evidencia su incapacidad para anticipar comportamiento que difieren de gran manera con sus datos históricos. Esta limitación, ya advertida en literatura crítica sobre modelización predictiva (27), recuerda que “todos los modelos son erróneos, algunos son útiles”, y obliga a mantener una interpretación crítica de sus resultados.

Otra línea de análisis reveló que países con proximidad geográfica como Ecuador, Perú y Colombia muestran tanto similitudes como divergencias notables en sus perfiles de mortalidad. Estas diferencias sugieren que, además de los factores geográficos o económicos, inciden variables contextuales como las políticas de salud, la capacidad de respuesta institucional, o la calidad de los sistemas de recolección de datos. Estudios como los de Cieza (28) han subrayado la importancia de estos factores institucionales en la variabilidad de los indicadores de salud entre países con características demográficas similares.

Finalmente, el estudio muestra que, si bien las enfermedades mentales no figuran entre las principales causas de muerte a escala global, su presencia en el top 10 —con la demencia en séptima posición— evidencia que se debe empezar a dar más importancia a dichas enfermedades.

Como línea de trabajo futuro, se sugiere complementar los modelos desarrollados con variables socioeconómicas, ambientales y de infraestructura sanitaria, lo que permitiría aumentar la capacidad explicativa y ofrecer diagnósticos más integrales. Asimismo, se

recomienda realizar estudios comparativos con los datos reales del GBD correspondientes a los años 2022, 2023 y 2024, a medida que estos estén disponibles, para validar empíricamente la calidad de las predicciones generadas.

## CONCLUSIONES

El presente estudio cumplió con sus objetivos al aplicar técnicas estadísticas multivariantes sobre los datos del Global Burden of Disease (GBD) para analizar la evolución y proyección de las causas de muerte a escala global y nacional entre 1990 y 2022. Se implementó un flujo analítico riguroso que incluyó limpieza, estandarización, reducción de dimensionalidad mediante Análisis de Componentes Principales (ACP), y modelado predictivo con redes neuronales LSTM y Random Forest.

Se logró reducir la dimensionalidad del conjunto de datos, agrupando 32 causas de muerte en cinco dimensiones latentes que reflejan patrones epidemiológicos comunes y albergaron el 84.9% de varianza acumulada. Esta estructuración permitió comprender la carga de mortalidad desde una perspectiva más integrada y sirvió como base para los modelos predictivos.

Las predicciones realizadas mediante Random Forest ( $R^2 = 0.99$ ; MAE = 512) mostraron un mejor desempeño que las obtenidas con LSTM ( $R^2 = 0.5246$ ; MAE = 0.2108), tanto en precisión como en capacidad explicativa, lo que valida su aplicación en contextos con alta heterogeneidad estructural. Se confirmó que analizar el conjunto completo de enfermedades, en lugar de abordarlas de manera aislada, permite capturar interacciones relevantes y aporta valor predictivo, respaldando enfoques médicos que reconocen la multicausalidad en las defunciones.

Entre los hallazgos más relevantes destaca la consolidación de las enfermedades no transmisibles (en particular las cardiovasculares y el cáncer) como principales causas de muerte en la mayoría de los países, incluyendo Ecuador. Estas enfermedades mostraron una persistencia sostenida en el tiempo, lo que refleja la limitada efectividad de las estrategias sanitarias actuales para su control.

El análisis comparativo entre países geográficamente cercanos (Ecuador, Perú y Colombia) evidenció tanto similitudes como diferencias en la intensidad y evolución de enfermedades específicas, atribuibles a condiciones estructurales y políticas sanitarias propias. Al comparar países con niveles socioeconómicos contrastantes (Níger, Ecuador y Estados Unidos), se observó que el estatus económico tiene mayor peso explicativo que la geografía, al determinar la prevalencia de enfermedades transmisibles en países de bajos ingresos y de enfermedades crónicas en países desarrollados.

En conjunto, los resultados obtenidos validan el uso de herramientas estadísticas multivariantes como una vía eficaz para identificar patrones complejos de mortalidad y generar insumos relevantes para la formulación de políticas públicas basadas en evidencia.

## REFERENCIAS BIBLIOGRÁFICAS

1. Hrzic R, Vogt T. The contribution of avoidable mortality to life expectancy differences and lifespan disparities in the European Union: a population-based study. *The Lancet Regional Health - Europe*. 2024 Nov 1;46.
2. OMS. Las diez causas principales de defunción [Internet]. 2024 [cited 2025 Apr 23]. Available from: <https://www.who.int/es/news-room/fact-sheets/detail/the-top-10-causes-of-death>
3. Espinosa Zúñiga JJ. Aplicación de algoritmos Random Forest y XGBoost en una base de solicitudes de tarjetas de crédito. *Ingeniería Investigación y Tecnología*. 2020 Jul 1;21(3):1–16.
4. Brauer M, Roth GA, Aravkin AY, Zheng P, Abate KH, Abate YH, et al. Global burden and strength of evidence for 88 risk factors in 204 countries and 811 subnational locations, 1990–2021: a systematic analysis for the Global Burden of Disease Study 2021. *The Lancet* [Internet]. 2024 May;403(10440):2162–203. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0140673624009334>
5. Carrillo Jesús, María del Rosario González. Iris: Codificador automático internacional de Causas de muerte. 2016.
6. Pita Ferreira P, Godinho Simões D, Pinto de Carvalho C, Duarte F, Fernandes E, Casaca Carvalho P, et al. Real-Time Classification of Causes of Death Using AI: Sensitivity Analysis. *JMIR AI*. 2023 Nov 22;2:e40965.
7. Falissard L, Morgand C, Roussel S, Imbaud C, Ghosn W, Bounebacher K, et al. A deep artificial neural network-based model for prediction of underlying cause of death from death certificates: Algorithm development and validation. *JMIR Med Inform*. 2020 Apr 1;8(4).
8. Liu J, Ning W, Zhang N, Zhu B, Mao Y. Estimation of the Global Disease Burden of Depression and Anxiety between 1990 and 2044: An Analysis of the Global Burden of Disease Study 2019. *Healthcare (Switzerland)*. 2024 Sep 1;12(17).
9. Wei S, Han Y, Liu M, Wang H, Lu Z, Cheng Y, et al. Burden of Pulmonary Arterial

Hypertension in China from 1990 to 2021, with Projections to 2050: An Analysis of Data from the Global Burden of Disease Study 2021 [Internet]. 2025. Available from: <https://www.researchsquare.com/article/rs-5904466/v1>

10. Bishop K, Balogun S, Eynstone-Hinkins J, Moran L, Martin M, Banks E, et al. Analysis of Multiple Causes of Death: A Review of Methods and Practices. *Epidemiology*. 2023 May 1;34(3):333–44.

11. Shen Z, Luo H. The impact of schistosomiasis on the Global Disease Burden: a systematic analysis based on the 2021 Global Burden of Disease study. *Parasite* [Internet]. 2025 Feb 21;32:12. Available from: <https://www.parasite-journal.org/10.1051/parasite/2025005>

12. Dai H, Alsalhe TA, Chalghaf N, Riccò M, Bragazzi NL, Wu J. The global burden of disease attributable to high body mass index in 195 countries and territories, 1990–2017: An analysis of the Global Burden of Disease Study. Vol. 17, *PLoS Medicine*. Public Library of Science; 2020.

13. Fernández L, Pérez M, Orduña JM, Alcaraz JM. A new dimensionality reduction technique based on the Wavelet Transform for cancer classification. *J Big Data*. 2025 Dec 1;12(1).

14. Shoaee S, Ghasemi E, Sofi-Mahmudi A, Shamsoddin E, Tovani-Palome MR, Roshani S, et al. Global, regional, and national burden and quality of care index (QCI) of oral disorders: a systematic analysis of the global burden of disease study 1990–2017. *BMC Oral Health*. 2024 Dec 1;24(1).

15. Azangou-Khyavy M, Ghasemi E, Rezaei N, Khanali J, Kolahi AA, Malekpour MR, et al. Global, regional, and national quality of care index of cervical and ovarian cancer: a systematic analysis for the global burden of disease study 1990–2019. *BMC Womens Health*. 2024 Dec 1;24(1).

16. Murray CJL, Lopez AD. Global mortality, disability, and the contribution of risk factors: Global Burden of Disease Study.

17. De Lucio J. Advanced estimation of regional growth using LSTM neural networks. *Investigaciones Regionales*. 2021;2021(49):45–64.



18. Foreman KJ, Marquez N, Dolgert A, Fukutaki K, Fullman N, McGaughey M, et al. Forecasting life expectancy, years of life lost, and all-cause and cause-specific mortality for 250 causes of death: reference and alternative scenarios for 2016–40 for 195 countries and territories. *The Lancet*. 2018 Nov 10;392(10159):2052–90.
19. IHME. Our World in Data [Internet]. 2024 [cited 2025 Apr 26]. Available from: <https://ourworldindata.org/grapher/burden-of-disease-by-cause?tab=table>
20. Brown MB, Forsythe AB. Robust Tests for the Equality of Variances. Vol. 69, Source: *Journal of the American Statistical Association*. 1974.
21. Arana C. UNIVERSIDAD DEL CEMA Buenos Aires Argentina Serie [Internet]. 2021. Available from: [www.cema.edu.ar/publicaciones/doc\\_trabajo.html](http://www.cema.edu.ar/publicaciones/doc_trabajo.html)
22. Patiño Pérez D, Silva Bustillos R, Munive Mora C, Botto-Tobar M. Predicción de Covid19 con el uso del Algoritmo Random Forest y Redes Neuronales Artificiales. *Ecuadorian Science Journal*. 2020 Sep 30;4(2):101–10.
23. Steinmetz JD, Culbreth GT, Haile LM, Rafferty Q, Lo J, Fukutaki KG, et al. Global, regional, and national burden of osteoarthritis, 1990-2020 and projections to 2050: a systematic analysis for the Global Burden of Disease Study 2021. *Lancet Rheumatol*. 2023 Sep 1;5(9):e508–22.
24. Díaz de León-Castañeda C. Revisión de temas fundamentales en sistemas de salud. 2018; Available from: <http://www.redalyc.org/articulo.oa?>
25. Naghavi M. Global, regional, and national burden of suicide mortality 1990 to 2016: systematic analysis for the Global Burden of Disease Study 2016. *BMJ*. 2019 Feb 6;364:l94.
26. Sánchez Iñigo. LA CAPACIDAD PREDICTIVA DE LAS REDES NEURONALES LSTM RESPECTO DEL BITCOIN. 2022.
27. Andersson C, Törnberg P. Wickedness and the anatomy of complexity. *Futures*. 2018 Jan 1;95:118–38.
28. Cieza A, Causey K, Kamenov K, Hanson SW, Chatterji S, Vos T. Global estimates of the



need for rehabilitation based on the Global Burden of Disease study 2019: a systematic analysis for the Global Burden of Disease Study 2019. The Lancet. 2020 Dec 19;396(10267):2006–17.