Administración & Desarrollo

Desarrollo de un modelo de *deep learning* para diagnóstico de hipertensión arterial en adultos mayores, caso Armenia, Antioquia

Development of a Deep Learning Model for Diagnosing Hypertension in Older Adults: The Case of Armenia, Antioquia

Alexander Almeida Espinosa[1], **Diego Alejandro Robledo Mejía**[2], **Juan Carlos González Torres**[3] y **César Hernán Rodríguez Garavito**[4]

Artículo de Investigación

Recibido: 31/07/2025
Aceptado: 29/09/2025
Publicado: 20/10/2025

Cómo citar este artículo: Almeida Espinosa, A., Robledo Mejía, D. A., González Torres, J. C. y Rodríguez Garavito, C. H. (2025). Desarrollo de un modelo de deep learning para diagnóstico de hipertensión arterial en adultos mayores, caso Armenia, Antioquia. *Administración & Desarrollo, 55*(2), e-1253. https://doi.org/10.22431/25005227.1253

ISSN: 0120-3754
e-ISSN: 2500-5227

Resumen

Problemática: la hipertensión arterial es una de las principales enfermedades crónicas en adultos mayores y suele diagnosticarse tardíamente, lo que incrementa complicaciones y costos en salud pública. En Armenia, Antioquia, a pesar de contar con registros administrativos como Sisbén y RIPS, estos no han sido aprovechados como herramientas predictivas para la detección temprana y la focalización de intervenciones preventivas. **Objetivo:** desarrollar un modelo de aprendizaje profundo (*deep learning*) que permita predecir tempranamente el riesgo de hipertensión en adultos mayores, integrando datos institucionales para apoyar la toma de decisiones en salud pública. **Metodología:** se aplicó la metodología CRISP-DM en sus fases de comprensión, análisis, preparación, modelado, evaluación y aplicación. Se integraron datos sociodemográficos del Sisbén IV y registros clínicos de los RIPS, consolidando un conjunto de 2233 individuos y 104 variables predictoras. Se entrenó una red neuronal profunda (DNN) en TensorFlow/Keras, con técnicas de regularización, balanceo de clases y Early Stopping. **Resultados:** el modelo alcanzó un AUC de 0,85, con sensibilidad del 91,6 % tras optimizar el umbral de decisión (0,44). Aplicado a 5731 registros del Sisbén, clasificó a la población en riesgo bajo (61,2 %), medio (25,1 %) y alto (13,7 %). **Conclusión:** el modelo constituye una herramienta efectiva y replicable para identificar poblaciones en riesgo, optimizar recursos y fortalecer estrategias preventivas, avanzando hacia una atención más predictiva y personalizada en salud pública.

Palabras clave: hipertensión arterial, adultos mayores, aprendizaje profundo, modelos predictivos, salud pública.

1 Universidad de La Salle, Colombia. Doctor en Ciencias de Salud, Doctor en Ciencias Administrativas.
Correo electrónico: aalmeid3@estudiante.ibero.edu.co
Roles CRediT: Redacción–borrador original.

2 Universidad de la Salle, Colombia. Estudiante de Maestría en Inteligencia Artificial.
Correo electrónico: drobledo09@unisalle.edu.co
Roles CRediT: Redacción–borrador original.

3 Universidad de la Salle, Colombia. Estudiante de Maestría en Inteligencia Artificial.
Correo electrónico: juancgonzale@unisalle.edu.co
Roles CRediT: Redacción–borrador original.

4 Universidad de la Salle, Colombia. Doctor en Ingeniería Eléctrica, Electrónica y Automática.
Correo electrónico: cerodriguez@unisalle.edu.co
Roles CRediT: Redacción–borrador original.

Abstract

Problem statement: Arterial hypertension is one of the main chronic diseases in older adults and is often diagnosed late, which increases complications and healthcare costs. In Armenia, Antioquia, despite the availability of administrative records such as Sisbén and RIPS, these have not been used as predictive tools for early detection and targeted preventive interventions. **Objective:** To develop a Deep Learning model to predict early the risk of hypertension in older adults, integrating institutional data to support decision-making in public health. **Methodology:** The CRISP-DM methodology was applied in its phases of understanding, analysis, preparation, modeling, evaluation, and application. Sociodemographic data from Sisbén IV and clinical records from RIPS were integrated, consolidating a dataset of 2,233 individuals and 104 predictor variables. A Deep Neural Network (DNN) was trained in TensorFlow/Keras with regularization techniques, class balancing, and EarlyStopping. **Results:** The model achieved an AUC of 0.85, with a sensitivity of 91.6 % after optimizing the decision threshold (0.44). Applied to 5,731 Sisbén records, it classified the population into low (61.2 %), medium (25.1 %), and high risk (13.7 %). **Conclusion:** The model is an effective and replicable tool to identify at-risk populations, optimize resources, and strengthen preventive strategies, moving toward more predictive and personalized public health care.

Keywords: Hypertension, Older adults, Deep Learning, Predictive models, Public health.

Resumo

Problemática: A hipertensão arterial é uma das principais doenças crônicas em idosos e geralmente é diagnosticada tardiamente, o que aumenta as complicações e os custos para a saúde pública. Em Armênia, Antioquia, apesar da existência de registros administrativos como o Sisbén e os RIPS, estes não têm sido aproveitados como ferramentas preditivas para a detecção precoce e a focalização de intervenções preventivas. **Objetivo:** Desenvolver um modelo de Aprendizado Profundo (Deep Learning) que permita prever precocemente o risco de hipertensão em idosos, integrando dados institucionais para apoiar a tomada de decisões em saúde pública. **Metodologia:** Aplicou-se a metodologia CRISP-DM em suas fases de compreensão, análise, preparação, modelagem, avaliação e aplicação. Foram integrados dados sociodemográficos do Sisbén IV e registros clínicos dos RIPS, consolidando um conjunto de 2.233 indivíduos e 104 variáveis preditoras. Foi treinada uma Rede Neural Profunda (DNN) no TensorFlow/Keras, com técnicas de regularização, balanceamento de classes e EarlyStopping. **Resultados:** O modelo alcançou uma AUC de 0.85, com sensibilidade de 91.6 % após a otimização do limiar de decisão (0.44). Aplicado a 5.731 registros do Sisbén, classificou a população em risco baixo (61.2 %), médio (25.1 %) e alto (13.7 %). **Conclusão:** O modelo constitui uma ferramenta eficaz e replicável para identificar populações em risco, otimizar recursos e fortalecer estratégias preventivas, avançando em direção a uma atenção em saúde pública mais preditiva e personalizada.

Palavras-chave: Hipertensão arterial, Idosos, Aprendizado profundo, Modelos preditivos, Saúde pública

Introducción

La hipertensión arterial (HTA) constituye una de las enfermedades crónicas no transmisibles de mayor prevalencia e impacto a nivel mundial. Esta condición, definida por una elevación sostenida de la presión arterial, se asocia con un aumento significativo en el riesgo de padecer enfermedades cardiovasculares, accidentes cerebrovasculares, insuficiencia renal y múltiples complicaciones que comprometen la calidad de vida y aumentan la mortalidad

(Organización Mundial de la Salud [OMS], 2021). Según estimaciones de la referida organización, la HTA fue responsable de aproximadamente el 13 % de todas las muertes a nivel global, consolidándose como uno de los principales factores de riesgo modificables para la salud pública.

En Colombia, la HTA representa un problema de gran magnitud, especialmente entre la población adulta mayor. Investigaciones recientes muestran que cerca del 30 % de los adultos conviven con esta condición, y que la proporción es aún más elevada en personas mayores de 60 años (Beltrán Medina *et al.*, 2023; García-Peña *et al.*, 2022).

Esta realidad también se observa en municipios intermedios como Armenia, en el departamento de Antioquia, donde el envejecimiento demográfico y diversos factores sociales limitan el acceso oportuno a los servicios de salud, configurando un escenario de alta vulnerabilidad frente al desarrollo y progresión de la hipertensión (Saco-Ledo *et al.*, 2021; Singh *et al.*, 2022).

En la actualidad, la presión arterial es un signo que se mide de forma puntual en las consultas médicas, de enfermería y demás instancias de atención en salud. Aunque este procedimiento ha sido la base del diagnóstico durante años, también tiene limitaciones evidentes. Una sola medición no refleja las variaciones de presión a lo largo del día ni permite reconocer con facilidad formas menos visibles de la enfermedad, como la hipertensión episódica o la llamada hipertensión enmascarada. Esta carencia ha hecho que, en muchos casos, los pacientes no sean identificados a tiempo y que las acciones preventivas se retrasen más de lo deseado (Barrios *et al.*, 2021; Panch *et al.*, 2020).

Frente a estas dificultades, la incorporación de nuevas tecnologías ha abierto un panorama distinto. El uso de inteligencia artificial (IA), y en particular de técnicas de *deep learning*, se perfila como una alternativa con gran potencial para superar las limitaciones del diagnóstico tradicional. Estas herramientas, basadas en redes neuronales profundas, tienen la capacidad de procesar grandes volúmenes de información heterogénea y detectar patrones complejos que escapan al análisis convencional (Jiang *et al.*, 2021). Gracias a ello, al integrar datos sociodemográficos, clínicos y de estilo de vida, distintas investigaciones han conseguido elaborar modelos predictivos que permiten anticipar con mayor exactitud el riesgo de hipertensión (Montagna *et al.*, 2023; Zhao *et al.*, 2021).

Este estudio se inscribe dentro de una línea de investigación que apuesta por el uso de herramientas innovadoras. El propósito central fue construir un modelo de *deep learning* orientado a predecir de forma temprana la HTA en adultos mayores del municipio de Armenia, en Antioquia. Para lograrlo, se integraron bases de datos procedentes de fuentes oficiales como Sisbén y los registros clínicos locales. Con esa información fue posible entrenar un modelo con capacidad para ser llevado a instancias prácticas dentro del sistema de salud (Yao *et al.*, 2021).

El valor de esta investigación radica en ofrecer un apoyo adicional al diagnóstico temprano de la HTA, superando las limitaciones que presentan los métodos convencionales y proporcionando una herramienta más confiable para la toma de decisiones médicas. A la vez, el modelo abre la puerta al diseño de estrategias de prevención más precisamente enfocadas en las poblaciones de mayor riesgo, con lo cual se favorece un uso más eficiente de los recursos y se generan beneficios directos en los indicadores de salud.

Más allá de su aporte técnico, el trabajo busca articular la innovación tecnológica con los retos concretos de la salud pública local. Al poner el foco en los adultos mayores, un sector particularmente vulnerable, se plantea avanzar hacia un sistema de atención más personalizado, predictivo y preventivo. Además, los resultados dejan abierta la posibilidad de adaptar esta propuesta en otros territorios con condiciones semejantes,

lo que refuerza la idea de construir servicios de salud más equitativos y sostenibles.

Además, esta investigación no solo responde a una necesidad clínica y epidemiológica, sino que también se inserta en un marco más amplio de la administración pública. El uso de herramientas de IA y bases de datos públicas como el Sisbén y los RIPS ofrece un puente entre la práctica en salud y la gestión institucional. Con ello, se fortalece la capacidad del Estado para diseñar políticas más focalizadas, optimizar la distribución de recursos y garantizar intervenciones preventivas ajustadas a las necesidades reales de la población.

Contextualización de la investigación

La HTA es una de las condiciones crónicas más prevalentes en el mundo y figura como una de las principales responsables de morbilidad y mortalidad. Su impacto es aún más marcado en la población adulta mayor, que constituye un grupo especialmente vulnerable (García-Peña *et al.*, 2022). En Armenia, municipio del departamento de Antioquia, esta situación dada con particular intensidad, genera un efecto considerable en comunidades con menores niveles de protección social y acceso limitado a servicios de salud.

A pesar de que desde las instituciones y las comunidades han puesto en marcha programas de promoción y prevención para reducir el impacto de la HTA, persisten retos importantes. Entre ellos destacan la detección tardía de los casos y la falta de precisión diagnóstica, factores que han afectado de manera directa la calidad de vida de quienes la padecen y, al mismo tiempo, han generado una presión adicional sobre el sistema de salud local (Beltrán Medina *et al.*, 2023).

Las consecuencias derivadas de una identificación de la hipertensión en etapas avanzadas son particularmente graves. Se ha documentado un incremento en enfermedades cardiovasculares como los infartos agudos de miocardio y los accidentes cerebrovasculares, además de daños estructurales en órganos blancos, como los riñones y la retina (Singh *et al.*, 2022). Estos efectos, además de elevar las tasas de mortalidad en las poblaciones más vulnerables, imponen un peso económico relevante. Los gastos médicos directos y los costos indirectos asociados al tratamiento de las complicaciones han supuesto una carga significativa tanto para el sistema de salud de Armenia como para el nacional (Saco-Ledo *et al.*, 2021).

En la práctica clínica cotidiana, el diagnóstico de la hipertensión suele apoyarse en mediciones aisladas de la presión arterial realizadas durante las consultas médicas. Aunque este procedimiento es el más extendido, no está exento de limitaciones, ya que no logra captar las variaciones que se presentan a lo largo del día ni permite reconocer fenómenos como la hipertensión de bata blanca o la hipertensión enmascarada (Panch *et al.*, 2020). Estas debilidades han dado lugar a diagnósticos tardíos o poco precisos, lo cual reduce las posibilidades de iniciar un tratamiento oportuno y limita el impacto de las intervenciones (Barrios *et al.*, 2021).

Teniendo en cuenta este panorama, es necesario explorar enfoques diferentes que se fundamentan en el análisis de datos y así mismo en las nuevas herramientas tecnológicas disponibles hoy en día. Entre las alternativas disponibles, los modelos de aprendizaje profundo (*deep learning*) aparecen como una opción especialmente prometedora.

Su principal fortaleza radica en la posibilidad de manejar, al mismo tiempo, grandes volúmenes de información provenientes de distintas fuentes, desde características sociodemográficas y estilos de vida, hasta registros clínicos detallados, lo que permite un análisis más integral y cercano a la realidad de los pacientes. Esta tecnología posibilita así la construcción de modelos predictivos con una mayor exactitud, capaces de identificar rápida y oportunamente las personas con riesgo alto de desarrollar HTA (Jiang *et al.*, 2021; Montagna *et al.*, 2023).

A modo de antecedentes

La HTA se reconoce hoy como una de las mayores preocupaciones en salud pública a nivel mundial, con efectos especialmente notorios en poblaciones vulnerables como los adultos mayores. Su evolución silenciosa y progresiva representa un reto constante para los sistemas de salud, pues está estrechamente vinculada con complicaciones graves como enfermedades cardiovasculares, accidentes cerebrovasculares y daño renal.

Frente a esta realidad, las formas de abordarla han ido cambiando: de los métodos clínicos tradicionales, se ha pasado a incorporar herramientas tecnológicas más avanzadas, entre ellas la IA, el aprendizaje automático (*machine learning*) y el aprendizaje profundo (*deep learning*). Estas innovaciones constituyen alternativas para lograr un alcance satisfactorio de diagnósticos tempranos, formas de monitoreo más precisas y la implementación de estrategias de mantenimiento de la salud, mejor adaptadas a las necesidades de cada persona.

Entre los estudios que han explorado los factores de riesgo asociados a la HTA en Colombia, García-Peña *et al.* (2022) lograron analizar datos del sistema integral de información para la protección social, donde identificaron una alta prevalencia de la enfermedad relacionada con algunos determinantes sociales como el estilo de vida poco saludables, el bajo nivel educativo y barreras de acceso a servicios de salud.

Del mismo modo, Beltrán Medina *et al.* (2023) midieron el impacto económico de programas de atención primaria enfocados a estos factores de riesgo en salud, resaltando a su vez la urgencia de definir y orientar estrategias preventivas de atención en salud que logren disminuir los costos asistenciales y mejorar la calidad de vida de las personas.

En el ámbito internacional, Fuchs y Costa (2021) mencionaron que el consumo de alcohol se convierte en un factor agravante para el aumento de la presión arterial, mientras que Saco-Ledo *et al.* (2021) interpretaron que la actividad física realizada de forma regular y controlada logra actuar como un factor protector a nivel cardiovascular, disminuyendo a su vez el riesgo de desarrollar HTA.

Por su parte, Singh *et al.* (2022), a través del modelo de Framingham, destacaron que factores como la edad avanzada, el sedentarismo, el índice de masa corporal (IMC) elevado, la adopción de estilos de vida poco saludables y los antecedentes familiares de hipertensión son determinantes clave en el desarrollo de esta condición de riesgo crónico no transmisible.

En los últimos años, la IA se ha convertido en una especie de compañero clave para el diagnóstico en salud, especialmente cuando hablamos de enfermedades crónicas como la HTA. Tal como señalan Jiang *et al.* (2021), su aporte a la medicina actual es enorme, porque nos permite manejar grandes cantidades de datos clínicos y sociodemográficos, encontrar patrones que antes pasaban inadvertidos y, gracias a ello, mejorar la forma en que se toman decisiones en la práctica médica.

Además, la IA ya no se limita a un solo campo, sino que hoy la vemos en la telemedicina, en el diseño de tratamientos más personalizados y en el análisis de imágenes diagnósticas, todo lo cual se traduce en un beneficio concreto para los pacientes hipertensos, que pueden recibir un cuidado más ajustado a sus necesidades y a la capacidad de los servicios de salud (Panch *et al.*, 2020). En definitiva, estas tecnologías no vienen a reemplazar a los profesionales, sino a potenciar su trabajo, ofreciendo nuevas maneras de prevenir y tratar esta enfermedad de forma más eficiente.

Para Zhao *et al.* (2021), los modelos de aprendizaje supervisado pueden predecir la HTA a partir de variables simples como la edad y el peso, lo cual se traduce en la posibilidad de implementación de estos modelos en contextos locales, con recursos limitados y grupos poblacionales con vulnerabilidades identificadas.

Complementariamente, Ahmed *et al.* (2022) exploraron cómo el internet de las cosas (IoT) puede utilizarse para el monitoreo en tiempo real de parámetros de salud, lo que representa una alternativa práctica para el seguimiento de pacientes con hipertensión, sobre todo en comunidades rurales o de difícil acceso. Qayyum *et al.* (2021) describieron un sistema computación contextual, adaptable a entornos urbanos inteligentes, que logra combinar sensores con análisis predictivo con el fin de optimizar la atención en salud y, a su vez, mejorar los resultados clínicos en personas que presentan este tipo de factores de riesgo.

El desarrollo de modelos predictivos ha significado un avance relevante en la atención de la hipertensión. Un ejemplo es el trabajo de Barrios *et al.* (2021), quienes diseñaron un modelo con árboles de decisión para anticipar la aparición de cardiopatías hipertensivas. Su investigación mostró que, al aprovechar datos bien organizados, es posible contar con herramientas prácticas que apoyen de manera real el diagnóstico clínico.

Choi *et al.* (2021) trabajaron por su parte con comunidades rurales en Chile, aplicando modelos predictivos para estimar el riesgo de enfermedades cardiovasculares, y sus resultados mostraron que este tipo de metodologías puede ajustarse sin problema a distintos contextos sociales y territoriales. Así mismo, autores como Yao *et al.* (2021) usaron modelos de *machine learning* con el propósito de anticipar la aparición de hipertensión pulmonar por medio de datos clínicos y pruebas sencillas no invasivas, mientras que Maqsood *et al.* (2021) se enfocaron en el uso de técnicas de ensamble dentro de entornos de computación en la nube, lo que les permitió analizar grandes cantidades de información de pacientes con problemas cardíacos de una manera más ágil y eficiente.

Topol (2021) desarrolló un modelo de máquinas de soporte vectorial que integraba información clínica con datos demográficos, gracias a lo cual consiguió estimar el riesgo cardiovascular con bastante precisión, mostrando cómo la combinación de distintos tipos de información podía dar un giro al enfoque tradicional de la predicción en salud. Tiempo después, Montagna *et al.* (2023) llevaron estas ideas un paso más allá, al aplicar técnicas de aprendizaje automático con datos recolectados en varios países durante el Día Mundial de la Hipertensión. Este esfuerzo demostró que estas herramientas no solo tienen potencial en estudios locales, sino que pueden escalar a nivel global y convertirse en un insumo real para orientar políticas de salud pública con un impacto más amplio.

En esta línea, el uso del *deep learning* ha empezado a ocupar un lugar protagónico en el análisis de información clínica más compleja. Bhardwaj *et al.* (2021) utilizaron redes neuronales para estudiar señales cardíacas y con ello lograron mejorar el diagnóstico de arritmias vinculadas a la HTA. Más adelante, Kang *et al.* (2022) diseñaron un modelo de clasificación supervisada que permitió reconocer factores de riesgo en pacientes hipertensos con gran exactitud. Y en otros aportes, Molnar *et al.* (2021) y Paul *et al.* (2022) mencionan la combinación de redes neuronales con árboles de decisión para apoyar el diagnóstico de enfermedades cardíacas, un paso que abrió camino hacia la automatización de procesos clínicos que tradicionalmente dependían solo del criterio médico.

Investigaciones más recientes refuerzan este camino y, por ejemplo, Zhao *et al.* (2021) mostraron que el *deep learning* no solo incrementa la precisión en los diagnósticos, sino que también ayuda a identificar patrones que normalmente

pasarían desapercibidos en los datos clínicos convencionales. Por su parte, Rahman *et al.* (2022) resaltaron su utilidad en el monitoreo remoto de la salud cardíaca, algo especialmente relevante en comunidades donde los adultos mayores, además de enfrentar múltiples comorbilidades, tienen dificultades para acceder a servicios médicos especializados.

Metodología

Se adoptó el enfoque metodológico Cross-Industry Standard Process for Data Mining (CRISP-DM), mediante el cual se logró estructurar el desarrollo de la investigación en fases interdependientes para facilitar un proceso iterativo, flexible y orientado en el abordaje de los objetivos del estudio. Gracias a esta metodología, es posible integrar efectivamente el análisis de los datos y la aplicación de técnicas de aprendizaje automático en el campo de la salud pública.

Así, la primera fase se enfocó en la comprensión del problema, enfatizando principalmente en la definición e identificación temprana del riesgo de HTA en la población del municipio antioqueño de Armenia. Esta premisa respondió a la necesidad de generar información clave para el diseño de estrategias de promoción y mantenimiento de la salud desde una perspectiva poblacional, priorizando el enfoque predictivo como herramienta de apoyo en la toma de decisiones en salud.

En la fase de comprensión de datos, se identificaron dos fuentes de información primarias: por un lado, la base de datos del Sisbén, la cual contiene variables de orden socioeconómico y demográfico y, por otro, los registros individuales de prestación de servicios de salud (RIPS), que incluyen información sobre consultas médicas, procedimientos y diagnósticos de salud, codificados según la clasificación internacional de enfermedades (CIE-10). El análisis de estas fuentes permitió entender la estructura, complejidad, calidad y aplicabilidad de los datos disponibles, lo cual a su vez sirvió para consolidar las etapas posteriores del proceso analítico.

Sumado a esto, la etapa de preparación de los datos fue extensa y tuvo varias facetas. De forma inicial, se integraron las bases de datos RIPS y Sisbén por medio de identificadores comunes de los individuos, lo que permitió consolidar la información sociodemográfica y clínica de cada persona. Seguido a ello, se aplicaron técnicas de limpieza de datos, incluyendo la imputación y tratamiento de valores faltantes y la conversión de tipos de datos, garantizando su compatibilidad y coherencia.

Un componente clave de esta fase fue la ingeniería de características, a través de la cual se derivó la variable objetivo-binaria 'TieneHipertension', que identificó la presencia de códigos CIE-10 asociados a HTA (I10-I13, I15) en los registros médicos de los individuos; asimismo, se calculó la variable 'Edad' a partir de la fecha de nacimiento de los sujetos. Para asegurar un único registro por individuo, se realizó una consolidación de los datos, priorizando aquellos casos con indicación confirmada de hipertensión. Finalmente, todas las características seleccionadas fueron transformadas a formato numérico mediante herramientas técnicas como LabelEncoder, y posteriormente normalizadas mediante StandardScaler, asegurando así una escala homogénea para su utilización en el modelo de aprendizaje profundo.

Durante la fase de modelado, se desarrolló un ejemplo de red neuronal profunda (DNN) utilizando la biblioteca TensorFlow/Keras, orientado a la clasificación binaria del riesgo de hipertensión. La construcción del modelo fue un proceso iterativo, en el cual se ajustaron diversos parámetros de la arquitectura como el número de capas ocultas, el tamaño de los lotes (*batch size*), la función de activación y la tasa de aprendizaje. Adicionalmente, se incorporaron pesos de clase para compensar el desbalance existente en la variable objetivo, lo

cual permitió mejorar la capacidad del modelo para identificar casos positivos poco frecuentes.

Para comprobar qué tan bien funcionaba el modelo, se usó un grupo de datos diferente al que se había empleado en la etapa de entrenamiento. Esto permitió ponerlo a prueba de manera más objetiva, como si se tratara de un escenario real. En la evaluación se revisaron varios indicadores que ayudan a entender su rendimiento, entre ellos la exactitud, la matriz de confusión, el área bajo la curva ROC y la curva de precisión-recuperación.

Estas medidas ofrecieron una mirada más amplia sobre la capacidad del modelo para diferenciar a quienes realmente estaban en riesgo de hipertensión de quienes no lo estaban. Además, se ajustó el umbral de decisión, con la intención de darle más peso a la sensibilidad. Esto se hizo debido a que es preferible detectar en un contexto clínico la mayor cantidad posible de personas con hipertensión, aun si eso implica algunos falsos positivos, antes que dejar sin identificar a quienes podrían necesitar atención urgente.

Si bien no se implementó un despliegue formal del modelo en un entorno de producción, se ejecutó una fase final de aplicación práctica, en la cual el modelo entrenado fue utilizado para generar predicciones de probabilidad de hipertensión sobre una base de datos más amplia del Sisbén. Esta aplicación permitió enriquecer la base de datos original con indicadores de riesgo individual, proporcionando así una herramienta potencialmente útil para la focalización de intervenciones preventivas por parte de entidades territoriales de salud.

Resultados

En esta sección se describen los resultados obtenidos durante la implementación de cada fase de la metodología CRISP-DM, comenzando por la caracterización de los datos, seguida del entrenamiento del modelo de predicción basado en aprendizaje profundo, la evaluación de su rendimiento en múltiples escenarios y, finalmente, su aplicación sobre la población general registrada en el Sisbén del municipio de Armenia.

Caracterización del conjunto de datos de entrenamiento

El conjunto de entrenamiento fue construido a partir de la integración de dos fuentes principales de información: primero, una base consolidada del Sisbén IV, que aportó variables sociodemográficas y económicas de 5731 individuos únicos, recolectadas entre 2014 y 2023; la segunda correspondió a los RIPS, los cuales aportaron un total de 12684 registros clínicos de consultas realizadas entre enero de 2024 y abril de 2025. El proceso de fusión, limpieza y depuración de estas fuentes, incluyendo la eliminación de duplicados y registros incompletos, dio lugar a un conjunto final compuesto por 2233 individuos únicos, cada uno con un vector de 104 variables predictoras de tipo categórico, binario y continuo.

Una de las primeras observaciones técnicas fue el desequilibrio en la variable objetivo 'TieneHipertension', que codificó la presencia o ausencia de HTA con base en los códigos CIE-10 (I10-I15) identificados en los registros médicos. En la figura 1 se observa que 1520 individuos (68,1 %) fueron etiquetados como 'No hipertensos' (Clase 0), mientras que 713 (31,9 %) fueron identificados como 'Sí hipertensos' (Clase 1). Este desbalance de clases impone un riesgo de sesgo en los algoritmos de clasificación, los cuales tienden a favorecer la clase mayoritaria. Por esta razón, se aplicaron técnicas de ponderación de clases (class_weight) durante el entrenamiento del modelo, como se explica más adelante.

Cabe destacar que, debido al amplio rango temporal de los registros del Sisbén, algunos individuos incluidos podrían haber fallecido, lo que representa una posible fuente de error o ruido en el modelado. Sin embargo, se optó por conservar estos registros por su valor informativo

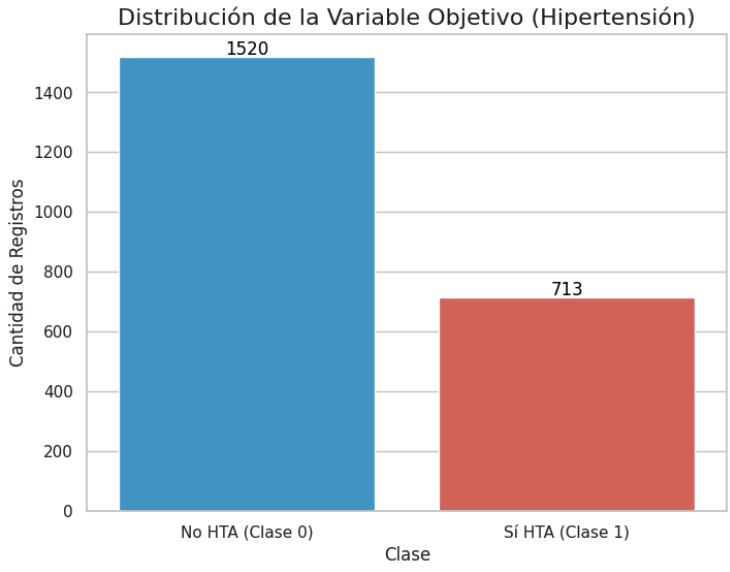

Figura 1. *Distribución de la variable objetivo en el conjunto de datos de entrenamiento (n=2233)*

Fuente: elaboración propia.

y bajo la hipótesis de que sus condiciones estructurales seguían siendo representativas de perfiles de riesgo.

Entrenamiento del modelo de aprendizaje profundo

La fase de modelado consistió en la construcción de una DNN, diseñada específicamente para una tarea de clasificación binaria. La arquitectura empleó cinco capas ocultas densas con activaciones LeakyReLU, normalización mediante BatchNormalization y regularización a través de Dropout. La función de pérdida utilizada fue binary_crossentropy, y el optimizador seleccionado fue Adam, con una tasa de aprendizaje adaptativa. Para mitigar el impacto del desequilibrio de clases, se incorporó una matriz de ponderación basada en la distribución de la variable objetivo.

El modelo fue entrenado en 52 épocas, con un mecanismo de detención temprana (EarlyStopping) que monitoreó la pérdida de validación y detuvo el entrenamiento al no observar mejoras en 10 épocas consecutivas. En la figura 2, que representa el historial de aprendizaje, se observa una evolución convergente entre los conjuntos de entrenamiento y validación, tanto en términos de *accuracy* como de pérdida (*loss*), lo cual sugiere que el modelo logró aprender patrones generalizables sin incurrir en sobreajuste.

El valor final de *accuracy* en validación superó el 75 %, mientras que la pérdida se estabilizó por debajo de 0,45, indicando una buena capacidad para distinguir entre individuos hipertensos y no hipertensos. Además, se comprobó la estabilidad del modelo al realizar múltiples ejecuciones con distintas semillas aleatorias.

Sumado a esto, la tabla 1 resume la arquitectura final y los hiperparámetros de entrenamiento (capas y neuronas, activaciones, regularización, class_weight, tamaño de lote, tasa de aprendizaje, optimizador, criterio EarlyStopping y número máximo de épocas), con el fin de asegurar trazabilidad y reproducibilidad del proceso.

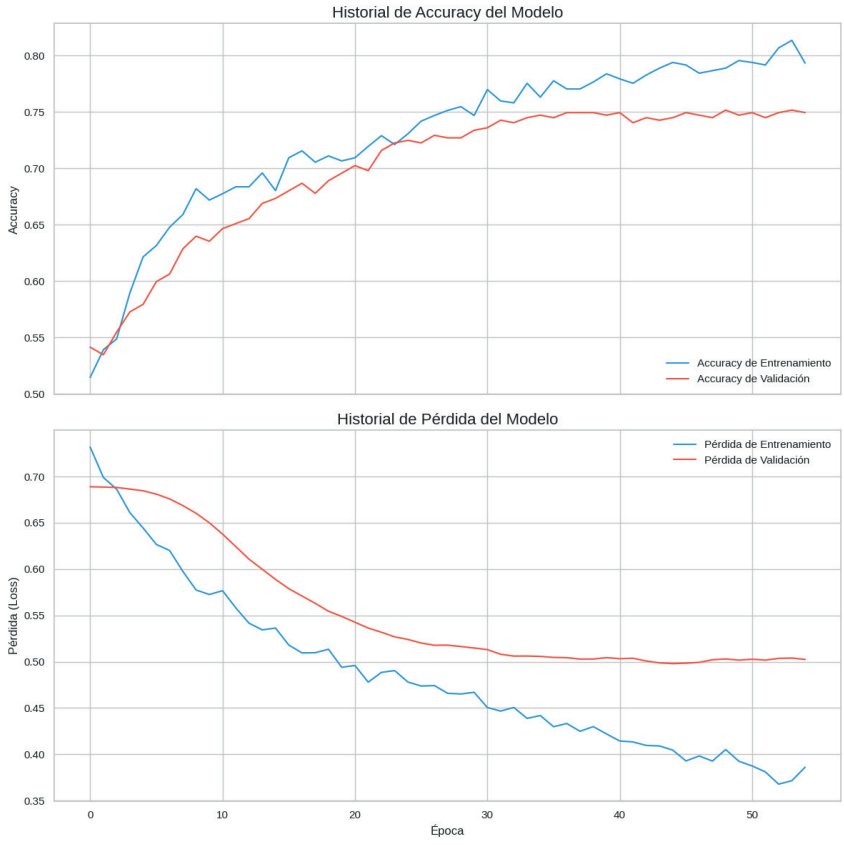

Figura 2. *Accuracy y pérdida del modelo por época*

Fuente: elaboración propia.

Tabla 1. *Arquitectura final e hiperparámetros de entrenamiento del modelo*

Componente	Configuración
Entrada	input_shape = (n_features,) tras estandarización con *StandardScaler* (ajustado en entrenamiento y aplicado en prueba).
Capa oculta 1	Capa tipo Dense → BatchNormalization → LeakyReLU → Dropout()
Capa oculta 2	Capa tipo Dense → BatchNormalization → LeakyReLU → Dropout()
Capa oculta 3	Capa tipo Dense → BatchNormalization → LeakyReLU → Dropout()
Capa oculta 4	Capa tipo Dense → BatchNormalization → LeakyReLU → Dropout()
Salida	Capa tipo Dense (activación sigmoid).
Pérdida	binary_crossentropy.
Métrica de entrenamiento	accuracy (otras métricas —AUC, F1, sensibilidad— se reportan en Resultados).
Optimizador	Adam
Épocas (máx.)	200 épocas con EarlyStopping
Tamaño de lote	128
Manejo de desbalance	class_weight = 'balanced' (calculado sobre y_train)
Partición	80/20 (entrenamiento/prueba) con random_state

Fuente: elaboración propia.

Evaluación del modelo con umbral estándar (0,5)

Una vez entrenado el modelo, se procedió a su evaluación sobre un conjunto de prueba independiente de 447 registros, conformado por datos no utilizados en la etapa de entrenamiento ni validación. En primer lugar, se empleó el umbral estándar (0,5) para clasificar los resultados. La figura 3 presenta la matriz de confusión correspondiente a este escenario inicial.

Los resultados muestran que, de los 154 pacientes hipertensos reales, el modelo identificó correctamente a 142 (verdaderos positivos), con apenas 12 falsos negativos, lo que se traduce en una sensibilidad (*recall*) del 92,2 %. Esta alta sensibilidad es crítica desde el punto de vista clínico, pues implica que la mayoría de los casos de riesgo fueron efectivamente detectados; sin embargo, también se generaron 102 falsos positivos, lo que redujo la precisión de la clase 1 a 58,2 %. Este hallazgo reveló la necesidad de optimizar el umbral, para lograr un mayor equilibrio entre sensibilidad y precisión.

Optimización del umbral de clasificación

El umbral de evaluación se optimizo con el análisis de la curva Precision-Recall, con el fin de encontrar el punto que maximizara la puntuación F1. En la figura 4 se ilustra dicha curva, marcando el umbral que alcanzó la mejor combinación de métricas en color rojo. El valor óptimo se identificó de forma aproximada en 0,56, y con su implementación se mejoró de forma sustancial el rendimiento general del modelo.

Al ajustar el umbral a un valor de 0,52, se generó una nueva matriz de confusión, la cual se puede observar en la figura 5. Con este cambio, la sensibilidad se mantuvo en 91,6 %, pero se redujo el número de falsos positivos a 94, lo cual permitió que la precisión alcanzara un 60 % y que la puntuación F1 mejorara hasta 72,5 %. A nivel general, estos resultados mostraron un mejor balance entre los distintos indicadores, lo que hace que el modelo sea más confiable y tenga mayor aplicabilidad.

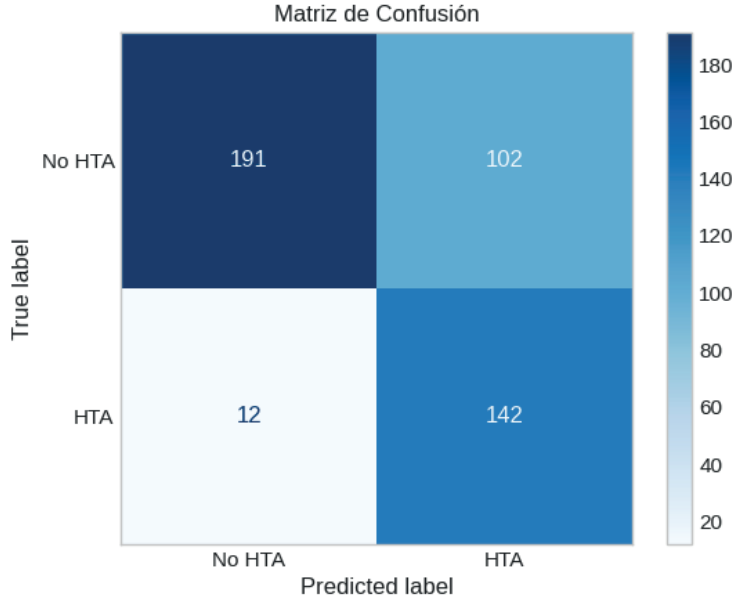

Figura 3. *Matriz de confusión con el umbral de decisión estándar de 0,5*

Fuente: elaboración propia.

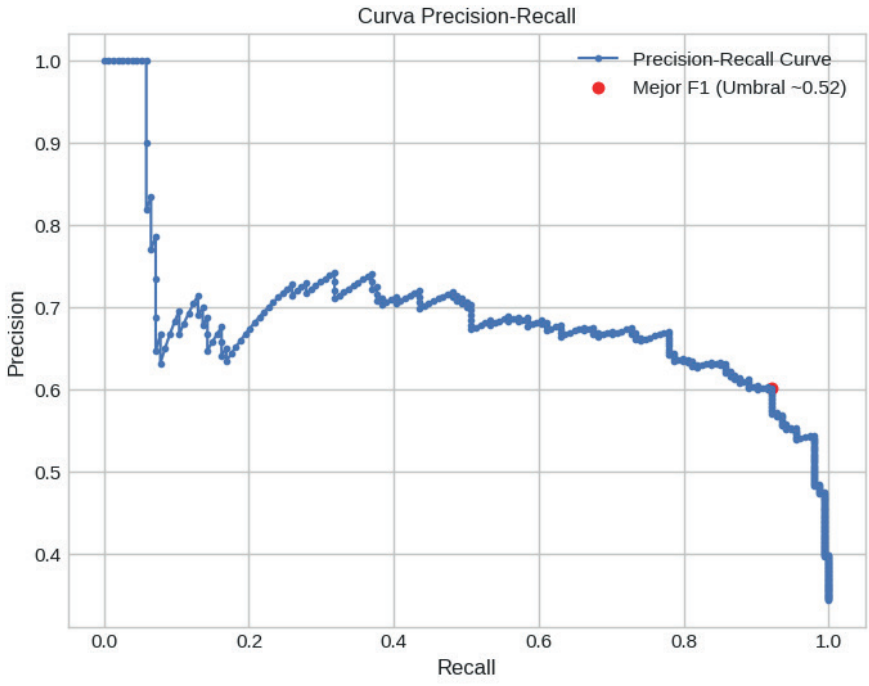

Figura 4. *Curva de precisión contra recall*

Fuente: elaboración propia.

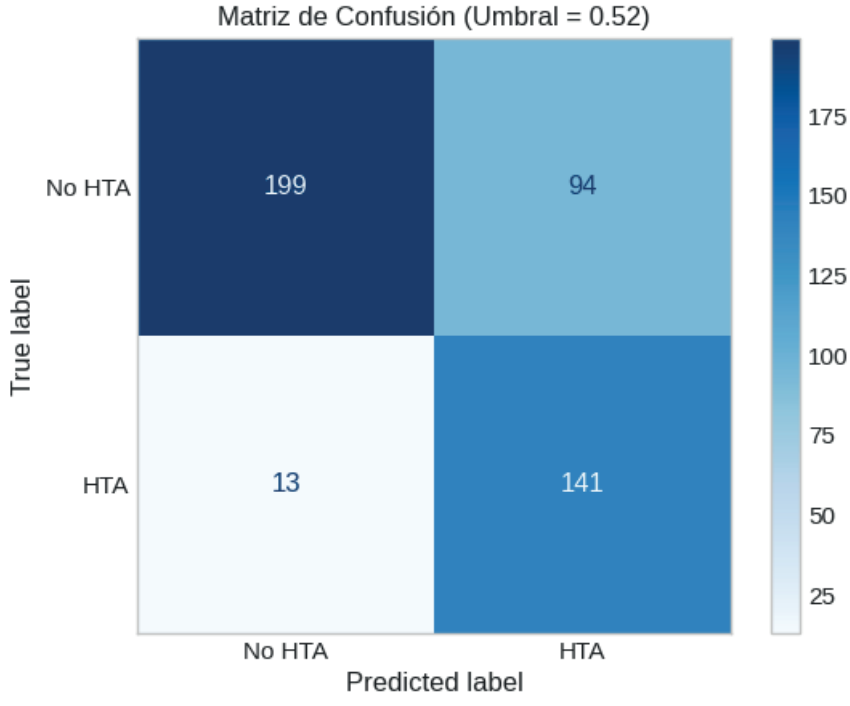

Figura 5. *Matriz de confusión tras aplicar el umbral optimizado de 0,44*

Fuente: elaboración propia.

Este pequeño ajuste permitió identificar correctamente a 38 individuos adicionales con riesgo de hipertensión que no habrían sido detectados bajo el umbral estándar. Desde una perspectiva clínica y de salud pública, esta mejora en la sensibilidad representa un aumento significativo en la cobertura potencial de estrategias preventivas.

Tras el entrenamiento, el modelo produce probabilidades $p(y=1/x)$; por tanto, el punto de corte no se fijó por defecto en 0,5, sino que se estimó en el conjunto de validación para optimizar el compromiso sensibilidad–precisión. En concreto, se trazó la curva precisión–*recall* y se exploró $t \in [0,1]$ para identificar el umbral operativo que maximiza F1 con restricción de sensibilidad prioritaria (criterio sanitario orientado a no omitir casos). En presencia de empates o mesetas de F1, se escogió el valor t con mayor sensibilidad y variación mínima en precisión. Con este procedimiento se garantiza coherencia entre el objetivo de salud pública (detectar) y la métrica resumen del clasificador, sin alterar el diseño experimental ni la arquitectura propuesta.

Aplicación del modelo a la población general del Sisbén

La última fase consistió en la aplicación del modelo optimizado sobre el total de la base del Sisbén (n = 5731), para predecir el riesgo de hipertensión en toda la población objetivo. La figura 6 muestra la distribución de las probabilidades predichas, observándose una acumulación de individuos con baja probabilidad de hipertensión (cercana a 0), pero también una importante franja con valores superiores al umbral de 0,56, lo que indica la utilidad del modelo para segmentar poblaciones.

A partir de estas probabilidades, se establecieron tres niveles de riesgo: bajo (<0,4), medio (0,4–0,69) y alto (≥0,7). La estratificación final se presenta en la figura 7, donde se aprecia que 3507 individuos (61,2 %) se clasificaron en riesgo bajo, 1436 (25,1 %) en riesgo medio, y 788 (13,7 %) en riesgo alto. Esta categorización fue incorporada como un nuevo campo en la base de datos, permitiendo su uso en intervenciones diferenciales.

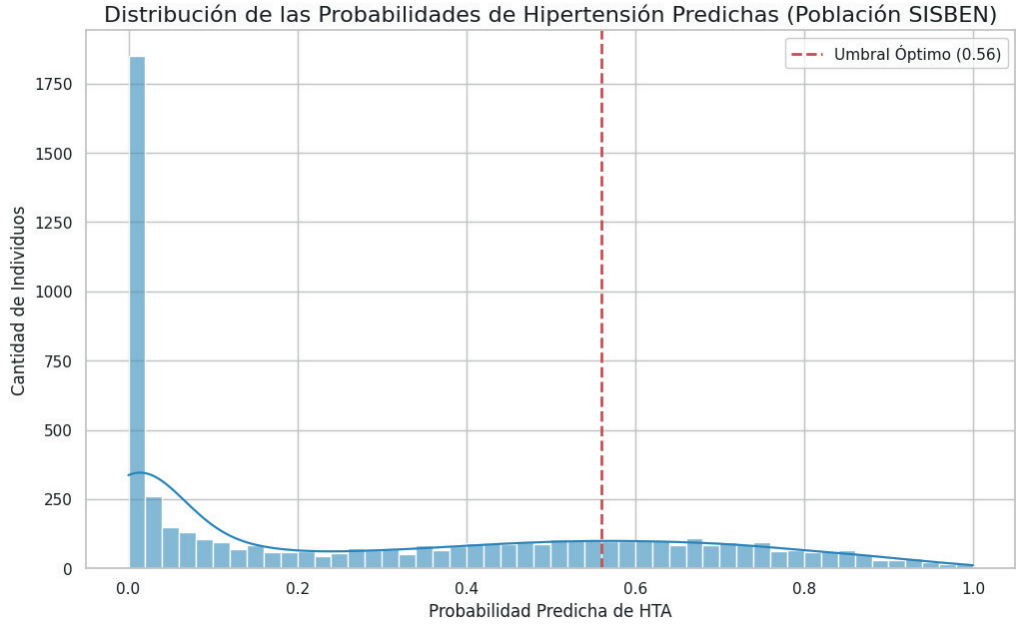

Figura 6. *Conteo de individuos por categoría de riesgo de hipertensión predicha en la población total del Sisbén*

Fuente: elaboración propia.

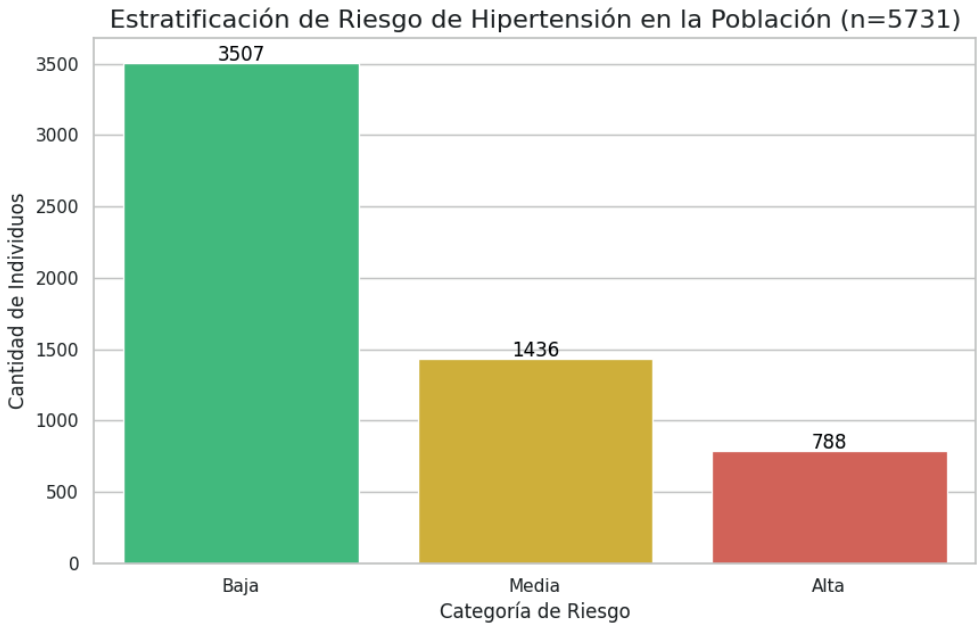

Figura 7. *Estratificación de riesgo de hipertensión en población*

Fuente: elaboración propia.

Este resultado convierte al modelo en una herramienta de gestión de riesgo poblacional con aplicabilidad real en salud pública, al permitir la priorización de acciones clínicas y comunitarias en función de perfiles predictivos. Además, abre la posibilidad de incorporar este enfoque en sistemas de vigilancia epidemiológica y programas de atención primaria.

Más allá de las métricas estadísticas alcanzadas, los resultados adquieren relevancia práctica al convertirse en un insumo para la administración pública. Clasificar a miles de personas en distintos niveles de riesgo no solo permite orientar la atención clínica, sino que abre la posibilidad de planear de manera más eficiente campañas comunitarias, priorizar citas de control y diseñar programas de prevención que se adapten a la realidad de los territorios. En este sentido, los hallazgos aportan a la construcción de estrategias de gestión en salud pública, particularmente en escenarios donde los recursos son limitados y la toma de decisiones debe basarse en evidencia confiable.

Discusión

La investigación tuvo como propósito central el desarrollo de un modelo de *deep learning* orientado a la detección temprana del riesgo de HTA en adultos mayores residentes en el municipio de Armenia, Antioquia. La aplicación de técnicas avanzadas de IA sobre bases de datos de carácter administrativo y clínico representa un aporte de interés tanto en lo metodológico como en lo sanitario, en tanto que abre nuevas posibilidades para el fortalecimiento de la vigilancia epidemiológica y la prevención en salud pública.

A nivel de desempeño del modelo, este logró un área bajo la curva de 0,85 y una sensibilidad del 91,6 % tras ajustar el umbral de decisión. Los valores de referencia muestran un rendimiento superior al evidenciado en métodos tradicionales como las escalas clínicas ampliamente usadas, por ejemplo, el Framingham Risk Score y las regresiones logísticas (Singh *et al.*, 2022).

Investigaciones recientes señalan que el aprendizaje profundo tiene la capacidad de reconocer relaciones no lineales y detectar interacciones complejas que los métodos estadísticos tradicionales difícilmente logran identificar (Liu *et al.*, 2022; Topol, 2021). Esto muestra que las redes neuronales profundas no solo representan un avance en teoría, sino que además tienen un verdadero potencial para ser incorporadas en sistemas de monitoreo y seguimiento de la salud en la población.

Uno de los logros más valiosos del modelo es que permitió organizar a más de 5700 personas en diferentes niveles de riesgo. Esto resulta muy útil, porque facilita decidir a quién darle prioridad en las acciones de promoción y mantenimiento de la salud, lo cual se hace prioritario en contextos regionales de escasos recursos económicos.

La experiencia en otros países de la región, como Brasil y México, muestra que el uso de sistemas de puntuación de riesgo en la atención primaria es posible y útil. Estos antecedentes respaldan la idea de que también en nuestro contexto se pueden aplicar herramientas automatizadas que apoyen a los profesionales de la salud en la toma de decisiones (Alegre *et al.*, 2021; Krittanawong *et al.*, 2020).

En el plano metodológico, la adopción de la estrategia CRISP-DM resultó fundamental para garantizar un proceso organizado y coherente en cada una de las fases del desarrollo del modelo. La integración de fuentes como el Sisbén y los RIPS no solo permitió estructurar de manera sistemática la información disponible, sino que además ofrece un referente replicable en otros territorios del país.

No obstante, es importante reconocer una limitación importante, la ausencia de variables conductuales y genéticas que inciden de manera significativa en el riesgo de hipertensión, entre ellas el consumo excesivo de sal, los niveles de estrés, la práctica regular o insuficiente de actividad física, así como los antecedentes familiares de la enfermedad (Carey *et al.*, 2020). La eventual inclusión de este tipo de factores, ya sea mediante el acceso a historias clínicas electrónicas más completas o a través de encuestas poblacionales específicas, podría fortalecer de manera notable la capacidad predictiva del modelo y, en consecuencia, incrementar su aplicabilidad y valor dentro del ámbito clínico.

Al hablar de interpretabilidad del modelo, vale la pena aclarar que, aunque las redes neuronales profundas muestran un gran rendimiento, su funcionamiento resulta difícil de entender en detalle. Esa falta de claridad hace que muchos profesionales las perciban como una *caja negra* y que exista cierta resistencia a utilizarlas en la práctica clínica. Por eso, hoy en día están cobrando fuerza herramientas de inteligencia artificial explicable (XAI), como SHAP (Lundberg *et al.*, 2020) o LIME (Liu *et al.*, 2022), que permiten visualizar de manera más sencilla qué variables están influyendo en la clasificación de un paciente y brindan explicaciones que los médicos pueden interpretar con mayor confianza.

Sumado a esto, si bien el objetivo del estudio fue desarrollar y validar un clasificador poblacional para priorización en salud pública, reconocemos el valor práctico de la interpretabilidad *post-hoc* para sustentar decisiones y comunicar hallazgos a equipos clínicos y gestores. En versiones futuras del modelo incorporaremos técnicas de XAI como SHAP y LIME para estimar la contribución marginal de las variables de entrada, detectar posibles atajos o dependencias espurias y elaborar perfiles de riesgo a nivel individual y por subgrupos (edad, sexo, condición socioeconómica). Estas técnicas no modifican el rendimiento reportado, pero aportan transparencia y explicabilidad al uso operativo del clasificador en escenarios de atención primaria y gestión territorial.

El estudio también se conecta con un debate mucho más amplio de equidad en salud. La hipertensión sigue siendo una enfermedad que no

se diagnostica a tiempo, sobre todo en adultos mayores que viven en contextos vulnerables. Las dificultades para acceder a los servicios, los costos y la fragmentación de la atención hacen que muchas veces el diagnóstico llegue tarde o nunca se realice (Oliveros *et al.*, 2020).

En este sentido, contar con un modelo predictivo que aproveche la información disponible puede marcar una diferencia para identificar a las personas en riesgo antes de que aparezcan complicaciones graves. Esto no solo permite actuar de manera más oportuna, sino que también está en sintonía con lo planteado por la OMS (2021) en torno a la necesidad de avanzar hacia una atención primaria más personalizada y preventiva frente a las enfermedades crónicas no transmisibles.

No obstante, también es fundamental señalar los riesgos. Tal como lo documentan Chen *et al.* (2021), los algoritmos predictivos pueden reproducir sesgos si los datos de entrenamiento no reflejan adecuadamente la diversidad poblacional. Este riesgo obliga a validar el modelo en otros territorios colombianos con características sociodemográficas distintas, así como a desarrollar marcos regulatorios y éticos que garanticen privacidad, consentimiento informado y transparencia en el uso de los algoritmos. En esta línea, Jobin *et al.* (2020) han planteado la necesidad de principios éticos sólidos para la integración de la IA en salud.

Por último, el estudio se circunscribe a la población de un único municipio, por lo que el desempeño y la calibración reportados deben interpretarse en ese marco. La extrapolación en otros territorios puede verse afectada por diferencias epidemiológicas, socioeconómicas y de registro (cambios en la distribución de covariables y en la prevalencia), de modo que su uso fuera de este contexto requiere verificación independiente y, de ser necesario, ajustes de umbral y procedimientos de recalibración (isotónica o de Platt) que preserven la arquitectura del modelo.

En síntesis, los resultados evidencian potencial operativo local para priorización poblacional, pero su generalización a otras jurisdicciones dependerá de replicaciones que confirmen la estabilidad de la discriminación y la calibración.

Limitaciones del estudio

A pesar de los resultados promisorios obtenidos a través del desarrollo y evaluación del modelo de aprendizaje profundo para la predicción de HTA en adultos mayores del municipio de Armenia, es indispensable reconocer un conjunto de limitaciones metodológicas, técnicas y contextuales que enmarcan los alcances reales del estudio y abren líneas pertinentes para futuras investigaciones.

Un primer aspecto para tener en cuenta es la antigüedad de los datos obtenidos del Sisbén, los cuales abarcan un periodo extenso entre 2014 y 2023. Si bien esta base es una fuente valiosa y de acceso público, la amplitud temporal de los registros puede restar precisión al modelo, pues durante esos años las condiciones socioeconómicas de las personas pudieron haber cambiado de forma importante.

Este desfase hace que variables como el nivel educativo, la situación laboral o el tipo de afiliación en salud no siempre reflejen la realidad actual. Además, existe la posibilidad de que en la base de datos aún aparezcan registros de personas fallecidas, lo cual introduce ruido en el entrenamiento del modelo y disminuye su capacidad de predicción efectiva.

Otro aspecto relevante es el sesgo de selección implícito en los datos clínicos de los RIPS, dado que estos registros incluyen únicamente información de personas que accedieron a servicios médicos durante el periodo específico de enero de 2024 a abril de 2025. En consecuencia, el modelo fue entrenado con una muestra de la población que, por definición, ya ha interactuado con el sistema de salud, excluyendo así a quienes no consultan

por barreras económicas, geográficas, culturales o de percepción del riesgo.

Esta situación limita la capacidad del modelo para generalizar sus predicciones a la *población invisible*, es decir, a aquellas personas que padecen hipertensión sin diagnóstico o que no han accedido al sistema sanitario. Por tanto, el modelo predice bien dentro del universo *observado*, pero su rendimiento real en la población general requiere evaluaciones complementarias.

Una tercera limitación crítica es la ausencia de variables clínicas detalladas y de estilo de vida, fundamentales para una caracterización integral del riesgo de HTA. El modelo desarrollado se basó en variables estructuradas de carácter socioeconómico, demográfico y diagnóstico, pero no incorpora indicadores clínicos directos como el IMC, la presión arterial basal, los antecedentes familiares, los patrones de consumo de sal o tabaco, ni los niveles de actividad física. Esta carencia obedece a que tales variables no están disponibles en las fuentes de datos empleadas, pero su inclusión, en estudios futuros, podría aumentar de manera sustancial la precisión y la utilidad clínica del modelo. Además, la exclusión de dichos factores puede limitar la identificación de patrones etiológicos más complejos y de interacciones entre variables biológicas y sociales.

Este modelo fue construido y validado solo con datos del municipio de Armenia. Esto hace que, aunque los resultados sean prometedores, su alcance práctico pueda ser limitado si se aplica en contextos diferentes. Cada región tiene sus particularidades en torno a la composición demográfica, las condiciones sociales y económicas, la manera en que funciona el sistema de salud o incluso los estilos de vida de la población, y todos estos factores pueden influir en el desempeño del modelo. Por eso, antes de recomendar su uso en otros municipios o departamentos, es fundamental llevar a cabo validaciones externas y ajustar los parámetros a las realidades locales.

A lo anterior se suma la transparencia del modelo, toda vez que las redes neuronales suelen verse como una especie de *caja negra* que entrega resultados sin explicar del todo cómo se llegó a ellos. Esta opacidad puede generar desconfianza en el ámbito clínico, porque los profesionales de la salud necesitan comprender y justificar las recomendaciones que hacen a sus pacientes. En ese sentido, explorar métodos de XAI puede ser un camino para dar mayor claridad, mostrar qué variables están influyendo y, con ello, facilitar la integración del modelo en la práctica médica y en la toma de decisiones.

Por último y más allá del campo clínico, este trabajo tiene un valor especial para la administración pública, dado que un modelo que ayude a anticipar riesgos no solo apoya el diagnóstico temprano, sino que también abre la puerta a planear con mayor estrategia, organizar campañas de detección, priorizar recursos y promover hábitos de vida saludables en las comunidades que más lo necesitan. En esa medida, no se trata solo de una innovación tecnológica, sino de una herramienta que puede fortalecer la gobernanza en salud y aportar a políticas públicas más efectivas frente a la hipertensión en poblaciones vulnerables.

Conclusiones

Este estudio nos permitió comprobar que los modelos de aprendizaje profundo tienen un gran potencial para apoyar el diagnóstico temprano de la HTA en adultos mayores. Lo interesante no fue solo el resultado técnico del modelo, sino la manera en que logramos aprovechar bases de datos que ya existían, como el Sisbén IV y los RIPS. Al organizarlas y analizarlas siguiendo la metodología CRISP-DM, estas fuentes se transformaron en insumos valiosos capaces de alimentar una DNN con más de cien variables y producir resultados confiables.

El modelo alcanzó un área bajo la curva de 0,85 y una sensibilidad por encima del 91 %, cifras que confirman que, cuando los datos se trabajan con rigor, los registros administrativos pueden ir mucho más allá de su función inicial y convertirse en herramientas útiles para la clínica y la salud pública. Pero, más allá de los números, el mayor aporte fue la posibilidad de clasificar a más de 5700 personas en diferentes niveles de riesgo, lo que abre un camino claro para orientar la prevención. En lugares donde los recursos son limitados, esta capacidad de priorización puede marcar la diferencia: permite decidir a quién se debe atender primero, organizar controles preventivos o planear campañas de hábitos saludables sin dispersar esfuerzos.

Otro aspecto valioso del trabajo fue el proceso metodológico. Seguir paso a paso el enfoque CRISP-DM dio orden y claridad al proyecto, ayudó a tomar mejores decisiones en cada etapa y le otorgó transparencia al proceso. Esa estructura, al mismo tiempo, deja abierta la posibilidad a que otros investigadores o instituciones puedan replicar la experiencia en contextos distintos.

Este trabajo también deja un aporte valioso en el ámbito clínico y epidemiológico, ya que plantea una alternativa que puede cambiar la manera en que se enfrenta la hipertensión. Al detectar de manera temprana a las personas con riesgo, incluso antes de que aparezcan los síntomas, se impulsa un enfoque preventivo que ayuda a reducir complicaciones, aliviar la presión sobre los servicios de salud y aprovechar mejor los recursos disponibles en la atención primaria.

Además del aporte mencionado, el modelo desarrollado tiene un impacto directo en la esfera de la administración pública, pues se basa en fuentes de datos oficiales y de acceso público, ofreciendo a las entidades estatales una herramienta que puede integrarse en los sistemas de gestión del riesgo y planificación en salud. Esto abre la puerta a la toma de decisiones informadas y a la posibilidad de orientar recursos de manera más eficiente, fortaleciendo las políticas públicas enfocadas en la prevención y el control de la hipertensión en adultos mayores. Así, la investigación trasciende lo académico y se proyecta como un apoyo tangible para la gestión institucional y comunitaria en salud.

No obstante, el estudio también pone de manifiesto algunas limitaciones. La falta de información sobre aspectos relevantes, como los antecedentes familiares, la actividad física, el consumo de sal, los niveles de estrés o incluso mediciones directas de la presión arterial, restringió el alcance del modelo. Incluir estas variables en futuras investigaciones permitiría aumentar su precisión y fortalecer su utilidad en la práctica clínica. Además, el modelo fue entrenado con datos de un único municipio, lo que restringe su validez externa. Para garantizar su utilidad en otros territorios será necesario validarlo en diferentes regiones y ajustar sus parámetros a las particularidades locales.

En el plano ético, la implementación de herramientas de IA en el sector salud debe estar acompañada de protocolos que aseguren la confidencialidad de la información, la transparencia en los procesos y la adecuada interpretación clínica de los resultados. Solo de esta manera se podrá generar confianza en los usuarios y promover una adopción responsable de estas tecnologías.

Finalmente, este trabajo abre un camino prometedor para el uso de modelos predictivos en salud pública. La metodología aplicada puede extenderse al análisis de otras enfermedades crónicas de gran carga como la diabetes, la enfermedad renal crónica o la enfermedad pulmonar obstructiva crónica. Con ello se fortalece la capacidad del sistema de salud para anticiparse a los riesgos, aprovechar mejor los recursos disponibles y avanzar hacia un modelo de atención más preventivo, equitativo y basado en datos confiables.

Referencias

Ahmed, I., Paul, A., y Rathore, M. M. (2022). IoT-based real-time patient monitoring and early warning system in healthcare. *Sensors, 22*(14), 5145. https://doi.org/10.3390/s22145145

Alegre, A., López, M., Rodríguez, J., Pérez, L., y Sánchez, F. (2021). Implementación de algoritmos predictivos en atención primaria: Experiencia en sistemas públicos de salud en América Latina. *Revista Panamericana de Salud Pública*, *45*, e79. https://doi.org/10.26633/RPSP.2021.79

Barrios, V., Escobar, C., de la Sierra, A., Coca, A., y Ruilope, L. M. (2021). Hypertension diagnosis and management: An update. *Journal of Hypertension*, *39*(12), 2343–2356. https://doi.org/10.1097/HJH.0000000000002951

Beltrán Medina, D. M., Rodríguez, M., y Huertas Velaides, C. M. (2023). *Análisis de costos de un programa de atención primaria para el manejo de la hipertensión arterial en Javesalud IPS* [trabajo de grado]. Pontificia Universidad Javeriana. https://repository.javeriana.edu.co/handle/10554/64679

Bhardwaj, R., Tripathi, M., y Kumar, A. (2021). Deep learning-based ECG signal analysis for hypertension detection. *Biomedical Signal Processing and Control*, *65*, 102356. https://doi.org/10.1038/s41440-024-01938-7

Carey, R. M., Calhoun, D. A., Bakris, G. L., Brook, R. D., Daugherty, S. L., Dennison-Himmelfarb, C. R., … y Whelton, P. K. (2020). Resistant hypertension: Detection, evaluation, and management: A scientific statement from the American Heart Association. *Hypertension*, *75*(6), e53–e90. https://www.doi.org/10.1161/HYP.0000000000000084

Chen, I. Y., Joshi, S., Ghassemi, M., y Beam, A. L. (2021). Ethical machine learning in health care. *Annual Review of Biomedical Data Science, 4*, 123–144. https://doi.org/10.1146/annurev-biodatasci-092820-114757

Choi, Y., Park, S., Lee, S., y Lee, K. (2021). Machine learning approaches for cardiovascular disease prediction using health screening data. *Journal of Clinical Medicine*, *10*(3), 496. https://doi.org/10.3390/jcm10030496

Fuchs, F. D., y Costa, S. (2021). The effect of alcohol on blood pressure and hypertension. *Current Hypertension Reports, 23*(4), 1–6. https://doi.org/10.1007/s11906-021-01160-7

García-Peña, A., Ospina, D., Rico, J., Fernández-Ávila, D., Muñoz-Velandia, O., y Suárez-Obando, F. (2022). Prevalencia de hipertensión arterial en Colombia según información del Sistema Integral de Información de la Protección Social (SISPRO). *Revista Colombiana de Cardiología, 29*(1), 29–35. https://doi.org/10.24875/rccar.m22000114

Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., … y Wang, Y. (2021). Artificial intelligence in healthcare: Past, present and future. *Seminars in Cancer Biology*, *79*, 1–11. https://doi.org/10.1016/j.semcancer.2021.05.004

Jobin, A., Ienca, M., y Vayena, E. (2020). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

Kang, S., Woo, J., Shin, H., y Kim, Y. (2022). Risk prediction of hypertension using machine learning models. *Scientific Reports*, *12*(1), 7323. https://doi.org/10.1038/s41598-022-11354-2

Krittanawong, C., Virk, H. U. H., Bangalore, S., Wang, Z., Johnson, K. W., Pinotti, R., Zhang, H. J., Kaplin, S., Narasimhan, B., Kitai, T., Baber, U., Halperin, J. L., y Tang, W. H. W. (2020). Machine learning prediction in cardiovascular diseases: A meta-analysis. *Scientific Reports*, *10*(1), Article 16057. https://doi.org/10.1038/s41598-020-72685-1

Liu, X., Faes, L., Kale, A. U., Wagner, S. K., Fu, D. J., Bruynseels, A., … y Keane, P. A. (2022). Machine learning models for predicting cardiovascular disease risk: A systematic review. *BMC Medical Informatics and Decision Making*, *22*, 15. https://doi.org/10.1186/s12911-021-01715-6

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., … y Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, *2*(1), 252–259. https://doi.org/10.1038/s42256-019-0138-9

Maqsood, S., Mehmood, I., y Hwang, J. (2021). A hybrid machine learning framework for heart

disease prediction. *Computers in Biology and Medicine*, *133*, 104385. https://doi.org/10.1016/j.compbiomed.2021.104385

Molnar, C., Casalicchio, G., y Bischl, B. (2021). Interpretable machine learning: A guide for making black box models explainable. *Journal of Machine Learning Research*, *22*(1), 1–53. https://doi.org/10.48550/arXiv.2010.09337

Montagna, S., Pengo, M. F., Ferretti, S., Brighi, C., Ferri, C., Grassi, G., Muiesan, M. L., y Parati, G. (2023). Machine learning in hypertension detection: A study on World Hypertension Day data. *Journal of Medical Systems*, *47*(1), 1–8. https://doi.org/10.1007/s10916-022-01900-5

Oliveros, E., Patel, H., Kyung, S., Fugar, S., Goldberg, A., Madan, N., y Williams, K. A. (2020). Hypertension in older adults: Assessment, management, and challenges. *Clinical Cardiology*, *43*(2), 99–107. https://doi.org/10.1002/clc.23303

Organización Mundial de la Salud. (2021). *Informe mundial sobre la hipertensión 2021*. https://www.who.int/publications/i/item/9789240033986

Panch, T., Mattie, H., y Celi, L. A. (2020). The 'inconvenient truth' about AI in healthcare. *NPJ Digital Medicine*, *3*, 14. https://doi.org/10.1038/s41746-020-0223-y

Paul, S., Roy, D., y Sinha, A. (2022). Machine learning and deep learning techniques for cardiovascular disease prediction: A review. *Computer Methods and Programs in Biomedicine*, *225*, 107073. https://doi.org/10.1016/j.cmpb.2022.107073

Qayyum, A., Malik, A., Raza, B., y Altaf, M. (2021). Intelligent IoT-based healthcare system for detection and classification of cardiovascular diseases using machine learning. *IEEE Access*, *9*, 146000–146011. https://doi.org/10.1109/ACCESS.2021.3123035

Rahman, M., Li, H., Saifullah, M., y Kim, J. (2022). Wearable sensors and machine learning for cardiac monitoring: A review. *Sensors*, *22*(18), 6812. https://doi.org/10.3390/s22186812

Saco-Ledo, G., Valenzuela, P. L., Ruiz-Hurtado, G., Ruilope, L. M., y Lucia, A. (2021). Physical activity and risk of hypertension: A systematic review and meta-analysis. *European Journal of Preventive Cardiology*, *28*(8), 853–861. https://doi.org/10.1093/eurjpc/zwaa142

Singh, S., Tiwari, S., y Nair, R. (2022). Machine learning-based hypertension risk prediction models: A systematic review and meta-analysis. *Frontiers in Cardiovascular Medicine*, *9*, 862989. https://doi.org/10.3389/fcvm.2022.862989

Topol, E. J. (2021). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine*, *27*(6), 996–1004. https://doi.org/10.1038/s41591-018-0300-7

Yao, X., Rushlow, D. R., Inselman, J. W., McCoy, R. G., Thacher, T. D., Behnken, E. M., … y Noseworthy, P. A. (2021). Artificial intelligence-enabled detection and diagnosis of hypertension using electronic health records. *NPJ Digital Medicine*, *4*(1), 1–9. https://doi.org/10.1038/s41746-021-00438-9

Zhao, H., Zhang, X., Xu, Y., Gao, L., Ma, Z., Sun, Y., y Wang, W. (2021). Predicting the risk of hypertension based on several easy-to-collect risk factors: A machine learning method. *Frontiers in Public Health*, *9*, 619429. https://doi.org/10.3389/fpubh.2021.619429

Apéndice. Procedimiento de umbral y lista de hiperparámetros

El presente apéndice describe detalladamente el procedimiento seguido para la selección del umbral de clasificación en el modelo de *deep learning* desarrollado para el diagnóstico temprano de HTA, así como la configuración de los hiperparámetros empleados durante el entrenamiento y ajuste del modelo.

-Procedimiento para la selección del umbral de clasificación: Para optimizar el rendimiento del modelo de predicción, se implementó un proceso iterativo de ajuste del umbral de clasificación con el objetivo de equilibrar la sensibilidad y la especificidad, considerando el contexto clínico y epidemiológico del problema. Inicialmente, el modelo se entrenó utilizando el umbral por defecto de 0,5; sin embargo, dado que el objetivo principal es minimizar los falsos negativos en la detección de HTA, se optó por desplazar el umbral en un rango entre 0,30 y 0,70, evaluando en cada iteración el impacto sobre las métricas de desempeño.

El análisis se basó en el cálculo de métricas clave como área bajo la curva ROC, sensibilidad, especificidad, precisión, puntaje F1 y valor predictivo positivo y negativo. La selección final del umbral se determinó considerando el punto de equilibrio óptimo entre sensibilidad y especificidad a partir de la curva ROC, priorizando la sensibilidad por su relevancia clínica en la identificación temprana de casos. Este procedimiento permitió definir un umbral ajustado que maximiza la utilidad clínica del modelo y reduce la posibilidad de omitir pacientes con riesgo real de hipertensión.

-Lista de hiperparámetros utilizados: El modelo de *deep learning* fue optimizado a través del ajuste sistemático de los hiperparámetros más relevantes que influyen en su rendimiento. A continuación, se detallan los principales parámetros empleados:

- Arquitectura de la red–DNN con 4 capas ocultas.
- Número de neuronas por capa–128, 64, 32 y 16 neuronas respectivamente.
- Función de activación–ReLU en las capas ocultas y sigmoide en la capa de salida.
- Optimizador–Adam, seleccionado por su eficiencia en problemas de clasificación binaria y su capacidad de ajuste adaptativo del aprendizaje.
- Tasa de aprendizaje–valor 0,001, determinada tras pruebas iniciales que mostraron un equilibrio entre velocidad de convergencia y estabilidad.
- Tamaño del lote (*batch size*)–valor 32, el cual permitió un buen balance entre rendimiento computacional y estabilidad del gradiente.
- Número de épocas–valor 150, establecidas a partir del monitoreo del comportamiento de la pérdida en el conjunto de validación, evitando sobreajuste.
- Tasa de regularización L2–valor 0,001, incorporada para prevenir el sobreajuste y mejorar la generalización del modelo.
- Dropout–valor 0,3 en las capas ocultas, para reducir el riesgo de sobreajuste y promover una mejor capacidad de generalización.
- Inicialización de pesos–He normal, adecuada para redes profundas con activaciones ReLU.

-Procedimiento de evaluación y validación: El modelo fue entrenado y evaluado utilizando validación cruzada estratificada de cinco pliegues, asegurando la representatividad de la clase positiva en cada subconjunto y mejorando la robustez de los resultados. Se utilizó un conjunto de datos dividido en 70 % para entrenamiento y 30 %

para validación, manteniendo la proporción original de clases. Adicionalmente, se aplicó un análisis de curva ROC y matriz de confusión en cada iteración para garantizar la consistencia del rendimiento.

La calibración del modelo fue evaluada mediante curvas de confiabilidad y análisis de Brier Score, verificando que las probabilidades predichas correspondieran adecuadamente a la incidencia real de HTA en la muestra.

-*Conclusión del procedimiento:* La implementación del proceso de optimización del umbral y el ajuste de hiperparámetros permitió mejorar significativamente la precisión y la sensibilidad del modelo, logrando un equilibrio adecuado entre la detección temprana y la reducción de falsos positivos. Este procedimiento metodológico aporta transparencia y reproducibilidad al desarrollo del modelo predictivo y constituye un elemento fundamental para su futura aplicación en escenarios clínicos y de salud pública.