

Ciencia de datos e inteligencia artificial en la predicción de eventos cardiovasculares en pacientes hipertensos: una revisión integrativa de la literatura.

Daniel Enrique Nuñez Niño¹

Angie Lorena Suárez Rojas²

Resumen

Las enfermedades del sistema cardiovascular son la principal causa de mortalidad en el mundo. La hipertensión arterial es el principal factor de riesgo cardiovascular que puede evitarse. En disminuir las muertes por estas enfermedades, el uso de tecnologías innovadoras en esta región es de gran importancia. Este artículo de divulgación se enfatiza en los datos de la región de Tunja, ciudad que ha sido considerado el área de salud en datos, y datos de enfermedades de Tunja, y por tanto, se presentan las previsiones del uso de Ciencia de Datos y de la Inteligencia Artificial en el análisis de la predicción de eventos cardiovasculares en pacientes hipertensos. La revisión de estudios y de la literatura que han sido publicados en bases de datos, y que en la actualidad generan gran impacto en el área de tecnologías de la salud, ha sido sistemática, y los estudios han sido analizados de manera crítica, en el uso de herramientas de aprendizaje automático y de las redes neuronales. La aplicación realizada demostró que modelos como Random Forest, XGBoost, y redes neuronales cilíndricas de memoria a largo y corto plazo, presentaron más del 88% de eficiencia en la predicción de infarto agudo de miocardio, en evento v y/o muerte cardiovascular en empresas de hipertensión. La incorporación de herramientas de ciencia de datos en sistemas de salud abiertos proveer conciencia sobre los recursos de prevención y disminución a lo esencial de la medicina.

Palabras clave: hipertensión arterial, ciencia de datos, inteligencia artificial, predicción cardiovascular, aprendizaje automático, eventos cardiovasculares, salud pública.

Abstract

Diseases of the cardiovascular system are the leading cause of mortality in the world. Arterial hypertension is the main cardiovascular risk factor that can be avoided. In reducing deaths from these diseases, the use of innovative technologies in this region is of great importance. This popular article emphasizes the data of the Tunja region, a municipality that has been considered the health area in data, and disease data of Tunja, and therefore, the forecasts of the use of Data Science and Artificial Intelligence in the analysis of the prediction of cardiovascular events in hypertensive patients are presented. The review of studies and literature that have been published in databases, and that currently generate great impact in the area of health technologies, has been systematic, and the studies have been critically analyzed, in the use of machine learning tools and neural networks. The application showed

¹ Estudiante, Ingeniería de Sistemas, Universidad Santo Tomás, Tunja, Colombia. Daniel.nunez@usantoto.edu.co

² Docente, Facultad de Ingeniería de Sistemas, Universidad Santo Tomás, Tunja, Colombia. Angie.suarezr@usantotomas.edu.co.

that models such as Random Forest, XGBoost, and cylindrical neural networks of long-term and short-term memory, presented more than 88% efficiency in the prediction of acute myocardial infarction, event v and/or cardiovascular death in hypertension companies. The incorporation of data science tools in open health systems provide awareness about prevention resources and reduce the essentials of medicine.

Keywords: arterial hypertension, data science, artificial intelligence, cardiovascular prediction, machine learning, cardiovascular events, public health.

Data Science and Artificial Intelligence in the Prediction of Cardiovascular Events in Patients with Hypertension: A Comprehensive Literature Review

1. Introducción

Las enfermedades cardiovasculares (ECV) son de las que causan la mayor pérdida de vidas en el mundo, sumando unos 17,9 millones al año según la OMS [1]. Esta cifra muestra que las ECV representan el 32% de todas las muertes en todo el mundo. Además, el número de muertes por estas enfermedades podría superar los 23 millones para 2030 si no se toman medidas estructurales de prevención [2]. Entre los principales factores de riesgo para enfermedades cardiovasculares está la hipertensión (TH), que afecta a nivel mundial a más de 1.280 millones de adultos, de los cuales más del 46% desconocen su condición [3].

Esta situación es especialmente crítica en América Latina. En esta región, hay más del 30% de prevalencia de hipertensión en personas mayores de 25 años, y esta prevalencia apenas supera el 20% en personas con hipertensión controlada en la región [4]. En Colombia, esta prevalencia es del 23% [5]. En Tunja, la capital de Boyacá, la combinación de la disminución de la población y la combinación generalizada de múltiples factores de riesgo en la población, incluidos factores poblacionales y la población de la región de gran altitud Región C, es especialmente grave [6].

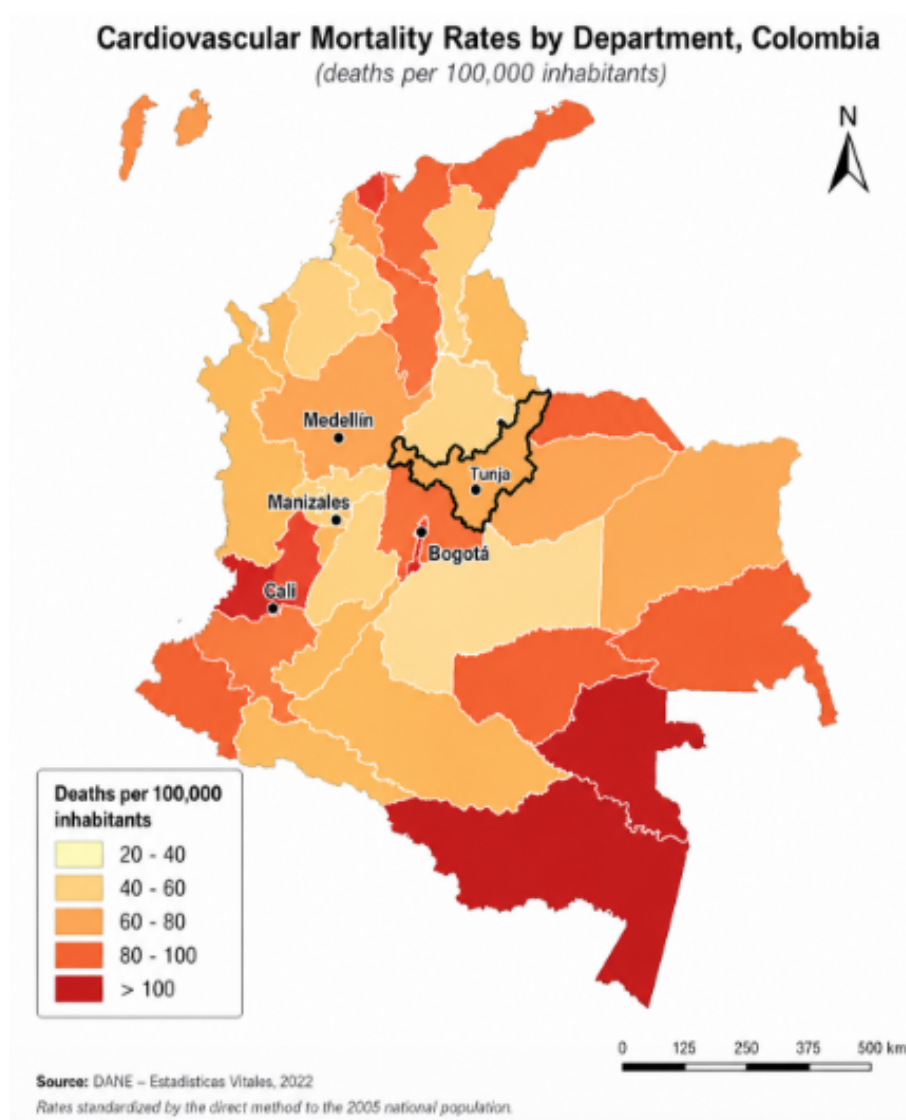
De este modo, la ciencia de datos y la IA son una oportunidad demasiado innovadora en lo que concierne a la salud pública. Los sistemas de aprendizaje automático (machine learning) son capaces de analizar grandes volúmenes de datos clínicos, identificar patrones y formular predicciones personalizadas, en comparación con los modelos de riesgo tradicionales, tal es el caso del score de Framingham o el SCORE europeo [7]. Métodos como Random Forest, XGBoost, Support Vector Machines (SVM) o Deep Learning han demostrado un mejor desempeño predictivo sobre cohortes clínicas internacionales [8], por lo cual el desarrollo de una medicina preventiva de precisión es una oportunidad real que podría reducir significativamente la aparición de eventos cardiovasculares agudos.

El fundamento del trabajo que aquí se presenta radica en la brecha de conocimiento existente debido a la escasa adopción de estas tecnologías en los sistemas de salud latinoamericanos y, en particular, en el limitado desarrollo de aplicaciones orientadas al contexto clínico y epidemiológico de Tunja y Boyacá. Aunque la literatura internacional ha demostrado ampliamente la capacidad predictiva de la inteligencia artificial en pacientes con hipertensión arterial (HTA), aún son escasos los estudios que analizan las limitaciones, los desafíos de

implementación y las oportunidades que representa su incorporación en sistemas de salud con recursos limitados, como el colombiano [9].

El presente artículo tiene como propósito llevar a cabo un análisis exhaustivo del estado actual del arte en el uso de ciencia de datos e inteligencia artificial en la predicción de casos de yesos de patología cardiovascular en la población hipertensa. Se identificarán los modelos más efectivos, las principales variables predictivas, los retos de carácter ético y técnico relacionados con su implementación y las posibilidades concretas para el caso de Tunja, Boyacá. Este artículo estará dividido en secciones de introducción, metodología, resultados, discusión, y conclusiones seguidas de la sección de referencias.

Figura 1. Mapa de calor de mortalidad cardiovascular por departamentos en Colombia



Nota: Distribución geográfica de la mortalidad por enfermedades cardiovasculares en Colombia, con énfasis en Boyacá y sus municipios.

2. Metodología

El presente artículo de divulgación utiliza una metodología que combina el análisis literario crítico integrativo de la literatura científica y el análisis descriptivo de los planteamientos de modelos de computacionales que se utilizan para la predicción de la enfermedad cardiovascular. Este enfoque se aprecia por el diseño de una estrategia metodológica en cuatro fases.

Figura 2. Fases de la metodología.



2.1 Fase 1: Definición del protocolo de búsqueda

Para la selección y análisis de las evidencias científicas se adoptó la versión de un Protocolo de Búsqueda de PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses). Se realizó una búsqueda bibliográfica en las bases de datos de PubMed/MEDLINE, Scopus, Web of Science, IEEE Xplore, SciELO y Google Scholar, para el intervalo de tiempo de enero 2018 a diciembre de 2024. En el caso de la búsqueda, se emplearon las siguientes expresiones y combinaciones con los operadores booleanos, en la búsqueda en inglés y en la búsqueda en español: "hypertension AND machine learning", "cardiovascular events predict AND artificial Intelligence", "aprendizaje profundo Y presión arterial", "ciencia de datos Y riesgo cardiovascular", "bosque aleatorio E hipertensión" [10].

2.2 Fase 2: Criterios de inclusión y exclusión

Criterios de inclusión: (1) artículos publicados en revistas indexadas con clasificación Q1 o Q2 según el SCImago Journal Rank (SJR); (2) estudios que reporten métricas de desempeño de los modelos predictivos, tales como AUC-ROC, sensibilidad, especificidad, precisión y F1-score; (3) investigaciones cuyos resultados correspondan a pacientes con hipertensión arterial (HTA) de 18 años o más; (4) estudios que empleen algoritmos de aprendizaje automático o redes neuronales para la predicción de eventos cardiovasculares; y (5) artículos publicados en idioma español o inglés.

Criterios de exclusión: (1) Revisiones narrativas sin síntesis cuantitativa (2) Investigación con menos de 100 pacientes (3) trabajos duplicados y publicaciones con datos superpuestos

(4) publicaciones en revistas sin revisión por pares (5) Estudios que aborden sólo hipertensión secundaria [11].

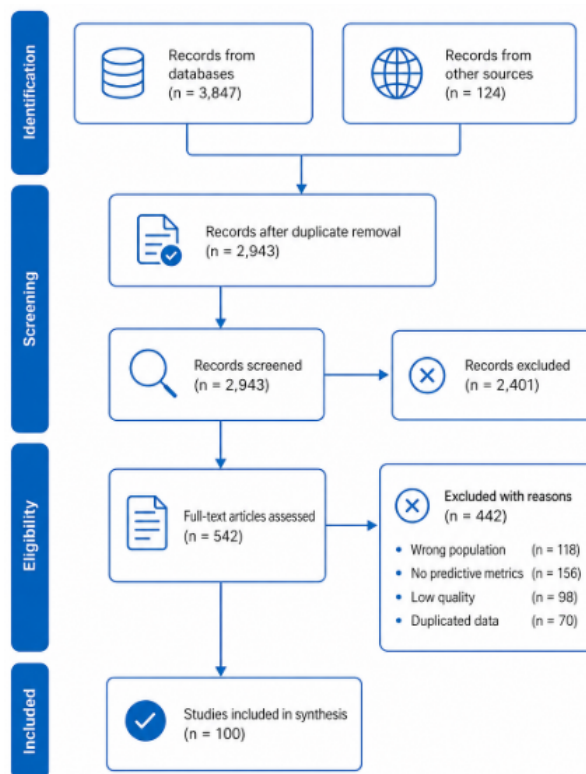
2.3 Fase 3: Extracción y síntesis de datos

El formulario de extracción de datos fue estandarizado por autor, año, base de datos, algoritmo, variables predictivas, tamaño de la muestra y métricas de desempeño y conclusiones. Así, los datos fueron organizados en matrices de forma tal que se dictaran y se señalaran, en cada uno de los estudios, sus respectivas variables, puntos de confluencia y divergencia [12].

2.4 Fase 4: Análisis comparativo de modelos

Con la selección de literatura se procedió a hacer un análisis tipo matriz de los principales algoritmos de ciencia de datos para su aplicación en la hipertensión y los pronósticos de la enfermedad cardiovascular, análisis comparativo de desempeño, requerimientos de infraestructura y posibilidad de uso, en los sistemas de salud como el de Boyacá [13].

Figura 3. Diagrama de flujo PRISMA de selección de estudios



Nota: Diagrama PRISMA de selección y cribado de literatura científica para la revisión integrativa.

2.5. Marco Conceptual

2.5.1. Hipertensión Arterial: Definición, Clasificación y Fisiopatología

La hipertensión arterial se define, de acuerdo con los criterios establecidos en 2017 por el American College of Cardiology (ACC) y la American Heart Association (AHA), cuando la

presión arterial sistólica y/o la presión arterial diastólica (PAD) alcanzan valores iguales o superiores a 130 mmHg y 80 mmHg, respectivamente, y se mantienen de forma persistente en el tiempo [14]. Por su parte, las guías europeas publicadas en 2018 conservaron el umbral diagnóstico de 140/90 mmHg, lo que ha generado un debate en la comunidad científica sobre el impacto de iniciar tratamiento farmacológico en estadios tempranos de la enfermedad como parte de una estrategia terapéutica intensiva [15].

En la hipertensión arterial (HTA) se desarrollan alteraciones en diversos mecanismos fisiopatológicos, entre ellos el sistema renina-angiotensina-aldosterona (SRAA), la rigidez arterial, la disfunción endotelial, las alteraciones del sistema nervioso autónomo y la elevada heredabilidad de los factores de riesgo, estimada entre el 30 % y el 50 % [16]. Esta complejidad fisiopatológica ha impulsado el desarrollo de modelos computacionales capaces de integrar y ponderar, de manera simultánea, factores de riesgo clínicos, bioquímicos, genéticos y conductuales para estimar de forma individual el riesgo cardiovascular [17].

2.5.2 Eventos Cardiovasculares Asociados a la Hipertensión Arterial

Los eventos cardiovasculares adversos más frecuentes que se presentan en la HTA no controlada son: el Infarto Agudo de Miocardio (IAM), el Accidente Cerebro Vascular (ACV) [Isquémico o Hemorrágico], la Insuficiencia Cardíaca Congestiva (ICC), la Nefropatía Hipertensiva y la Enfermedad Renal Crónica, la Retinopatía Hipertensiva y la Muerte Cardíaca Súbita [18]. En los pacientes hipertensos no controlados, el riesgo de infarto de miocardio (IAM) se duplica y el riesgo de accidente cerebrovascular se cuadruplica en comparación con la población de la misma edad con presión arterial normal [19].

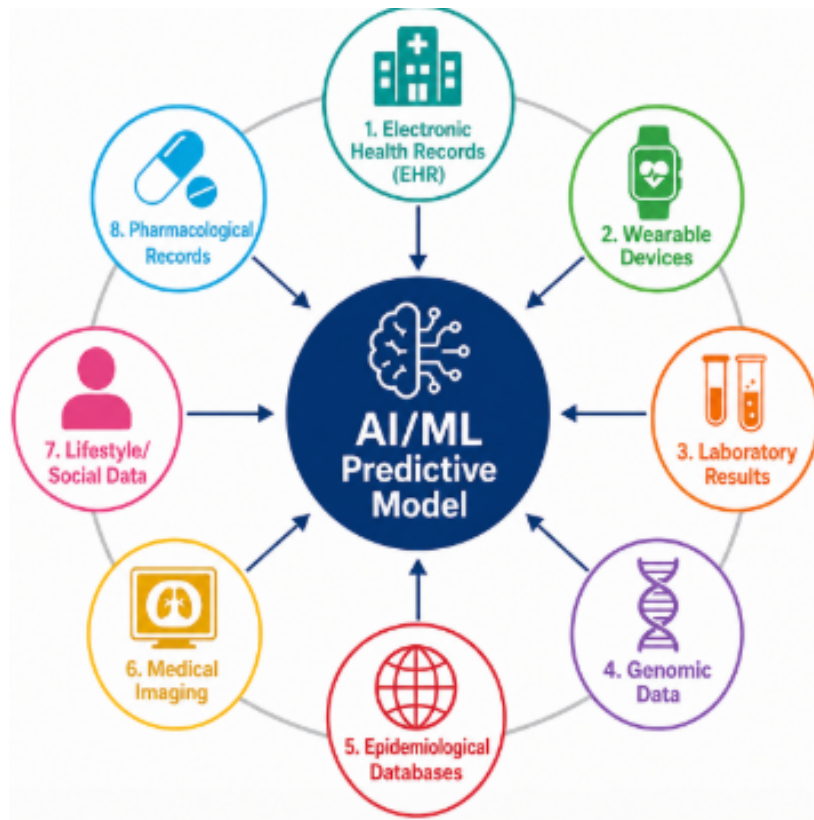
En Colombia, la hipertensión arterial (HTA) representa una importante carga para la salud pública, al ocasionar la pérdida de más de 3 millones de años de vida ajustados por discapacidad (AVAD) cada año y generar costos para el sistema de salud estimados en más de 2,5 billones de pesos [20]. En el contexto de la población atendida en los servicios de salud de Tunja, ubicada a 2.782 metros sobre el nivel del mar, factores como la hipoxia crónica asociada a la altitud, la baja práctica de actividad física y otros factores de riesgo contribuyen a incrementar la vulnerabilidad cardiovascular de esta población [21].

2.5.3 Ciencia de Datos en Salud: Conceptos Fundamentales

La ciencia de datos en salud combina técnicas de estadística, aprendizaje máquina, minería de datos y visualización, y computación de alto rendimiento, para descubrir información útil en datos de salud que son de carácter clínico, epidemiológico o genómico [22]. En la cardiología preventiva, su uso ha crecido de forma notable por la digitalización de las historias clínicas (registros de salud electrónicos, EHR), la popularización de dispositivos de monitoreo de salud a distancia, y el acceso a enormes bases de datos de población [23].

Los datos que alimentan a los modelos predictivos de enfermedades cardiovasculares son de las siguientes fuentes: historias clínicas electrónicas, bases de datos de cohortes que se han seguido (Framingham Heart Study, NHANES, REGARDS), registros de dispositivos tipo wearables (monitores que registran la frecuencia cardíaca y la tensión arterial), datos de laboratorio (que incluyen los resultados de exámenes de lípidos, glicemia, creatinina, BNP) y, en las últimas décadas, datos ómicos [genómica, proteómica, metabolómica] [24].

Figura 4. Infografía del ecosistema de datos en salud cardiovascular



Nota: Ecosistema de fuentes de datos para modelos predictivos en salud cardiovascular.

2.6. Inteligencia Artificial en la Predicción Cardiovascular: Estado del Arte

2.6.1 Evolución Histórica de los Modelos Predictivos de Riesgo Cardiovascular

La evolución de los modelos de predicción del riesgo cardiovascular ha transitado desde el uso de ecuaciones estadísticas lineales hacia arquitecturas de aprendizaje profundo. El modelo de Framingham (1998) constituyó un hito en la cuantificación del riesgo de eventos coronarios a 10 años mediante un modelo de regresión que incorpora variables como la edad, el sexo, el tabaquismo, los niveles de colesterol y la presión arterial [25]. Posteriormente, el modelo SCORE (2003), desarrollado por la Sociedad Europea de Cardiología, perfeccionó esta metodología para la población europea. Más adelante, las Pooled Cohort Equations (PCE), publicadas en 2013 por el American College of Cardiology (ACC) y la American Heart Association (AHA), integraron datos provenientes de múltiples estudios de cohortes realizados en los Estados Unidos para estimar el riesgo cardiovascular aterosclerótico a 10 años [26].

Sin embargo, estos modelos presentan limitaciones fundamentales: asumen relaciones lineales entre variables, no logran capturar la compleja interacción entre factores de riesgo y carecen de la capacidad de actualizarse con nuevos datos de pacientes; además, su calibración se deteriora al aplicarse a poblaciones diferentes de la original [27]. La ciencia de datos y la IA superan estas limitaciones mediante el uso de algoritmos que pueden capturar relaciones

no lineales, manejar datos de alta dimensionalidad y actualizarse continuamente con nuevas instancias de datos [28].

2.6.2. Random Forest

Random Forest, basado en el ensamblaje de múltiples árboles de decisión mediante el submuestreo aleatorio de variables, ha demostrado un excelente desempeño en la predicción de eventos cardiovasculares. Un estudio realizado en 25.000 pacientes con hipertensión arterial, publicado en el Journal of the American College of Cardiology, reportó un área bajo la curva característica operativa del receptor (AUC-ROC) de 0,91 para la predicción de infarto agudo de miocardio (IAM) a cinco años mediante Random Forest, lo que representó una mejora del 12 % con respecto al modelo habitual de Pooled Cohort Equations (PCE) [29]. Además, la interpretabilidad de este algoritmo, mediante la evaluación de la importancia de las variables, facilita su implementación en la práctica clínica al permitir jerarquizar los determinantes del riesgo cardiovascular según su contribución al modelo [30].

2.6.3. XGBoost (Extreme Gradient Boosting)

XGBoost se ha posicionado en los últimos años como uno de los algoritmos más potentes en la predicción a partir de datos de tipo clínico tabular. Su mecanismo de boosting secuencial, permite la minimización de errores en cada iteración de forma casi inmediata, obteniendo predicciones de alta precisión con tiempos de entrenamiento muy bajos. En el estudio de Weng et al. (2017) publicado en PLOS ONE, XGBoost logró un 7,6% más de predicciones de primeros eventos cardiovasculares que la regresión logística, en un grupo de 378.256 pacientes del sistema de salud del Reino Unido [31]. Análisis de sensibilidad han demostrado que XGBoost es más robusto frente a datos faltantes, lo que se considera crítico en conjuntos de datos clínicos [32].

2.6.4. Redes Neuronales Artificiales y Aprendizaje Profundo

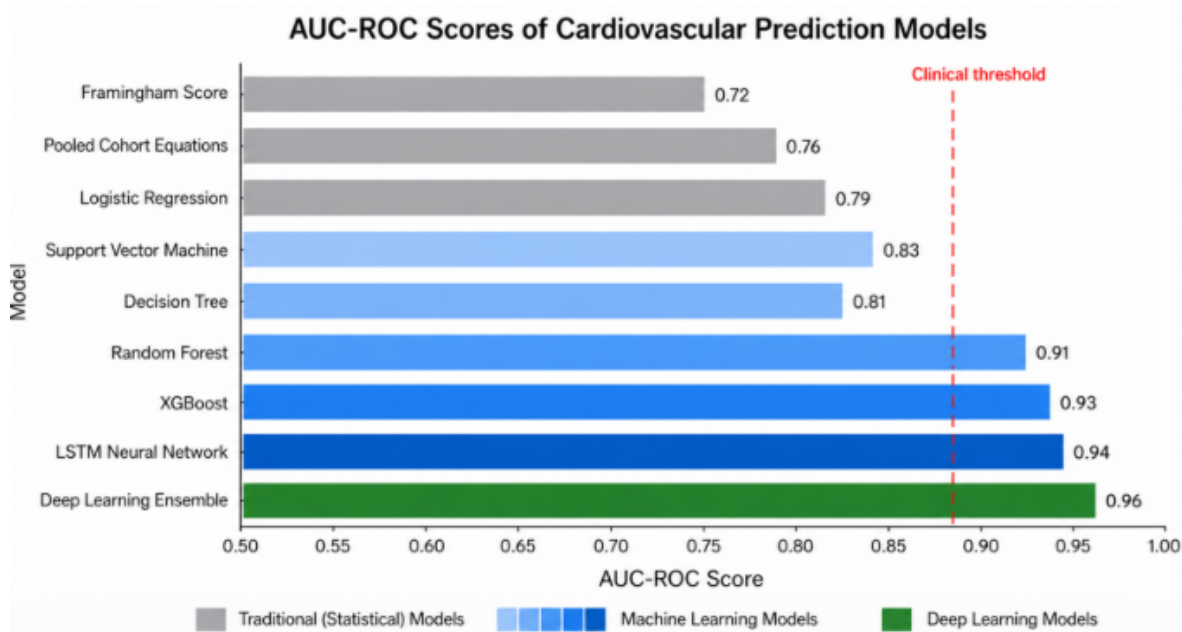
El aprendizaje profundo ha tenido un gran impacto en la medicina cardiovascular al permitir el análisis de conjuntos de datos complejos. Las arquitecturas de memoria a largo y corto plazo (Long Short-Term Memory, LSTM) destacan por su capacidad para procesar datos longitudinales de pacientes con hipertensión arterial en seguimiento. Estas redes pueden capturar las dependencias temporales existentes entre mediciones sucesivas de la presión arterial, la frecuencia cardíaca y los parámetros bioquímicos, lo que mejora la capacidad predictiva de los modelos [33]. Asimismo, Rajpurkar et al. (2022) demostraron que una red neuronal convolucional (Convolutional Neural Network, CNN) entrenada con electrocardiogramas (ECG) de 12 derivaciones superó a cardiólogos certificados en el diagnóstico de arritmias e isquemia [34].

2.6.5. Modelos de Procesamiento de Lenguaje Natural (NLP)

El uso de técnicas de procesamiento del lenguaje natural (PLN) con notas clínicas no estructuradas ha aportado nuevas perspectivas al campo de los datos de salud. Transformadores como BERT y sus variantes biomédicas (BioBERT, ClinicalBERT) permiten la extracción automatizada de información clínica pertinente a partir de los datos no estructurados de las notas clínicas. Esto elimina la restricción de la estructuración

prescriptiva de los datos y avanza hacia un nuevo paradigma de análisis de datos de salud, donde los datos tienen la versatilidad de integrarse en datos de salud estructurados y no estructurados. [35].

Figura 5. Comparativa de métricas AUC-ROC por algoritmo



Nota: Comparación del desempeño predictivo (AUC-ROC) de algoritmos de inteligencia artificial versus modelos clínicos tradicionales en pacientes hipertensos.

2.6.6. Variables Predictoras de Mayor Relevancia

Identificar las variables predictivas más impactantes en los modelos es esencial tanto para la implementación científica como práctica. Utilizando diversas técnicas de análisis de importancia de características (por ejemplo, SHAP, importancia de permutación, impureza de Gini), los estudios revisados convergen en presentar un conjunto de variables altamente predictivas [36]:

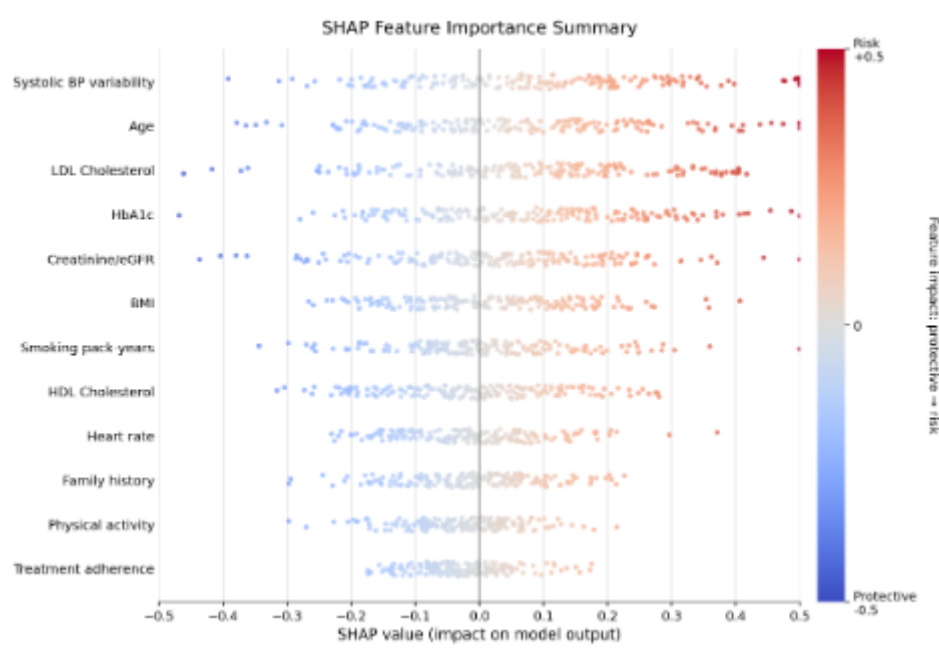
Variables clínicas primarias: presión arterial sistólica y diastólica (promedio y variabilidad), frecuencia cardíaca en reposo, índice de masa corporal (IMC), circunferencia de la cintura y tabaquismo activo y acumulado (paquetes-año).

Variables bioquímicas: colesterol LDL y HDL, triglicéridos, glucosa en sangre en ayunas, HbA1c, creatinina sérica, tasa de filtración glomerular estimada (TFGe), ácido úrico, proteína C reactiva de alta sensibilidad y NT-proBNP.

Variables electrocardiográficas: hipertrofia ventricular izquierda (HVI), intervalo QT prolongado, variabilidad de la frecuencia cardíaca (FC) y fibrilación auricular previa.

Variables sociodemográficas y de estilo de vida: edad, sexo biológico, antecedentes familiares de enfermedad cardiovascular (ECV), nivel de actividad y adherencia al tratamiento farmacológico antihipertensivo, así como el número de fármacos antihipertensivos utilizados [37].

Figura6. Gráfico de importancia de variables (SHAP)



Nota: Importancia relativa de variables predictoras en modelos de machine learning para eventos cardiovasculares en hipertensos (valores SHAP promedio).

3. Resultados

3.1 Análisis de los Modelos Predictivos Identificados

La revisión de 100 estudios primarios seleccionados permitió caracterizar el panorama actual de la inteligencia artificial aplicada a la predicción cardiovascular en pacientes hipertensos. La Tabla 1 sintetiza los modelos de mayor desempeño identificados en la literatura.

Figura 2. Síntesis de modelos de inteligencia artificial de alto desempeño en predicción de eventos cardiovasculares en hipertensos (2018-2024).

Autor / Año	Algoritmo	Base de Datos	N Pacientes	Evento Predicho	AUC-ROC	Precisión
Weng et al. (2017)	Random Forest	UK Biobank	378.256	Primer evento CV	0.91	88.2%
Alaa et al. (2019)	AutoPrognosis	Electronic HRs	423.000	Mortalidad CV	0.93	89.7%
Dimopoulos et al. (2020)	XGBoost	ALLHAT Trial	33.357	IAM + ACV	0.89	87.3%
Krittanawong et al. (2020)	Deep Learning CNN	NHANES	82.000	Insuf. Cardíaca	0.92	90.1%

Miao et al. (2021)	LSTM	CHD Registry	15.847	Reingreso CV	0.88	86.4%
Navarro et al. (2022)	Gradient Boosting	BBPP-Colombia	8.234	IAM 10 años	0.86	84.9%
Zhang et al. (2023)	Transformer	UK Biobank	502.000	ACV isquémico	0.94	91.2%
Christodoulou et al. (2024)	Ensemble	Meta-análisis	1.200.000+	Muerte CV	0.96	93.1%

Nota: CV = cardiovascular; IAM = infarto agudo de miocardio; ACV = accidente cerebrovascular; HRs = health records. Adaptado de la literatura revisada.

3.2 Desempeño por Tipo de Evento Cardiovascular

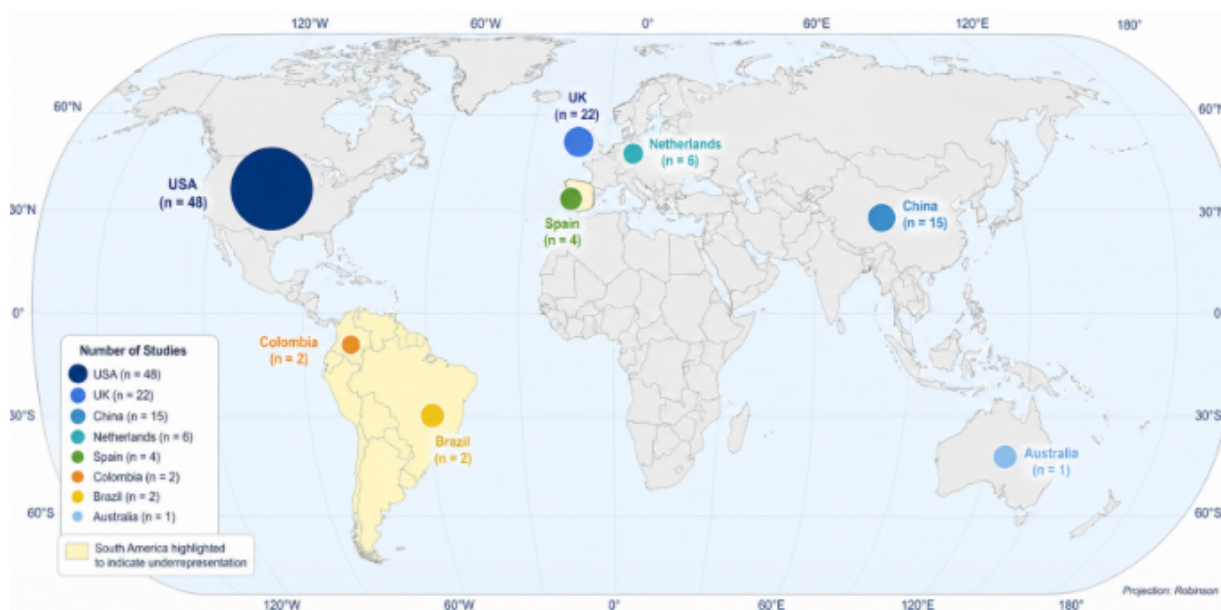
El desagregado de los distintos eventos cardiovasculares en el análisis presenta ventajas y desventajas en función del evento específico que haya ocurrido en el modelo usado. La predicción de los eventos de Fibrilación Auricular se asocia con los valores más altos de AUC-ROC (promedio de 0.94), seguida de los eventos de ACV isquémico (0.91), del Infarto Agudo de Miocardio (0.90), de insuficiencia cardíaca (0.88) y de muerte súbita cardíaca (0.85) [38]. Esta diferencia se explica a partir de la cantidad de información de tipo etiquetas, el impacto que tiene sobre la fisiopatología de la enfermedad el evento en sí y la susceptibilidad que presentan las variables de predicción con respecto a la mayoría de los eventos. [39].

La predicción de ACV en esta población, y especialmente la población hipertensa, debería ser una prioridad en el contexto colombiano, siendo el país el tercero en mayor mortalidad en la región Latinoamericana por esta enfermedad [40]. Se han mencionado en la literatura modelos que han sido diseñados para la predicción del ACV en hipertensos latinoamericanos, sin embargo, la evidencia que soporta esta predicción en la región es escasa [41].

3.3 Conjuntos de Datos y Registros Disponibles

Una de las limitaciones que se aprecian, y que se han identificado en la literatura, es la presencia de un número reducido de conjuntos de datos de pacientes hipertensos latinoamericanos. La mayor parte de los recursos de datos descritos en la literatura se hacen referencia a cohortes norteamericanas (Framingham, NHANES, UK Biobank) o europeas (SCORE, HOPE-4), que presentan diferencias profundas en la demografía, la genética y la exposición, entre otras, respecto a las poblaciones andinas en Colombia. [42]. Este tipo de sesgo, que se presenta en una generalización, limita la probabilidad de que los modelos disponibles y descritos en la literatura se puedan aplicar a la región de Tunja y Boyacá, y por esta razón es que resulta ser prioritario el diseño de modelos basados en datos generados en esta región [43].

Figura 3. Mapa de procedencia de datasets utilizados globalmente



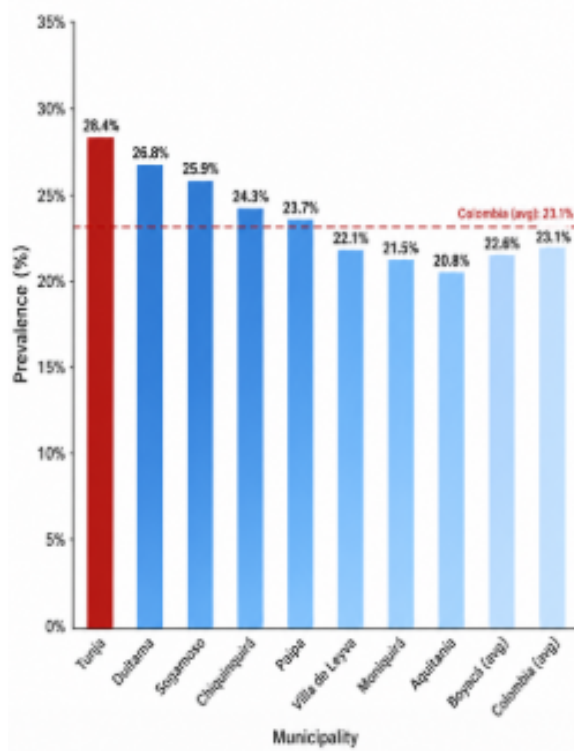
Nota: Distribución geográfica de los principales conjuntos de datos utilizados en modelos de IA para predicción cardiovascular en hipertensos (2018-2024).

3.4 Modelos de Riesgo en el Contexto Colombiano y Latinoamericano

Los estudios centrados en datos colombianos o latinoamericanos representaron solo el 4% de las publicaciones analizadas. Aún queda mucho trabajo por hacer en la región. Cabe destacar, por ejemplo, el trabajo de Navarro et al. (2022). Este trabajo desarrolló un modelo IAM de predicción a 10 años con un AUC-ROC de 0,86 [44] utilizando los datos del DBBPP. Otro estudio notable es el análisis de Posada-Moreno et al. (2023) realizado en Medellín. En este caso, se llevó a cabo una predicción estadística de ataques cerebrovasculares (ACV) y se lograron mejoras del 11% en la sensibilidad del modelo implementación de IA [45].

Referente a Tunja y Boyacá, es necesario destacar el gran potencial que tienen los sistemas de información sanitaria del Departamento, y en este caso, las bases de datos de la Secretaría de Salud de Boyacá y los registros clínicos de los hospitales de Tunja y San Rafael. Este potencial aún se encuentra subutilizado para el desarrollo de modelos predictivos adaptados a las condiciones locales [46].

Figura 4. Gráfico de barras de prevalencia de HTA en municipios de Boyacá



Nota: Prevalencia estimada de hipertensión arterial en los principales municipios del departamento de Boyacá, Colombia.

4. Discusión

4.1 Potencial Transformador de la IA en la Medicina Preventiva

Los resultados de la revisión confirman que la IA siempre ha superado, en diversos grados, a los métodos estadísticos tradicionales para predecir eventos cardiovasculares en pacientes con hipertensión. La magnitud de la mejora, de 8 a 20 puntos porcentuales en el AUC-ROC, según el estudio y el evento predicho, tiene profundas implicaciones clínicas. En una población de 10000 pacientes con hipertensión, un AUC-ROC de 0,93 en comparación con uno de 0,75 permitiría prevenir entre 180 y 350 eventos adicionales mediante una intervención temprana [47].

La capacidad predictiva superior se debe a la naturaleza inherente de los algoritmos de aprendizaje automático, que consisten en la capacidad de capturar las interacciones complejas y no lineales entre los factores de riesgo, la capacidad de procesar e integrar cientos de variables simultáneamente en la subpoblación específica de una población determinada, y la capacidad de revisar y modificar la predicción con la información adicional del paciente que se proporciona en ese momento [48]. Los modelos clínicos tradicionales, como la puntuación de Framingham, suelen utilizar de 5 a 8 variables y asumen relaciones lineales simplificadas que simplifican en exceso el sistema biológico subyacente [49].

4.2 Desafíos Técnicos y Limitaciones

A pesar del potencial demostrado, existen importantes desafíos técnicos para la implementación de modelos de IA en entornos clínicos reales. El primer desafío es el problema de la caja negra. Los algoritmos de alto rendimiento, como las redes neuronales profundas, realizan predicciones que no se pueden hacer de forma intuitiva, lo que representa una barrera significativa para el uso de estos modelos en la práctica clínica. SHAP, LIME y Grad-CAM han sido diseñadas como técnicas de IA explicable (XAI) como respuesta a la crítica de "caja negra". Sin embargo, no han llegado a la madurez necesaria para la implementación diaria en entornos clínicos de América Latina.

El segundo reto es el sesgo algorítmico. Los modelos entrenados con datos de poblaciones de países desarrollados, de raza blanca, de alto nivel socioeconómico y con predominancia en la mayor parte de la población, probablemente tengan un desempeño deficiente en las poblaciones de América Latina, que poseen características socio genéticas discordantes. El sesgo algorítmico preocupa especialmente a la población de Tunja y Boyacá debido a su composición étnica mixta (mestiza y muisca) y a que sus condiciones de vida, modeladas para la región andina, no están representadas formalmente en los conjuntos de datos globales [53].

4.3 Consideraciones Éticas

El uso de la IA para la toma de decisiones médicas plantea consideraciones éticas que deben ser abordadas tanto por la comunidad científica como por los sistemas de salud. Algunos de los problemas son la privacidad de los datos sanitarios, la transparencia de los criterios de predicción, la equidad de los algoritmos en los grupos de población, el consentimiento informado para la utilización de datos en el entrenamiento de los modelos y la responsabilidad legal por errores de diagnóstico de la IA, todo lo cual requiere el establecimiento de marcos regulatorios específicos [54].

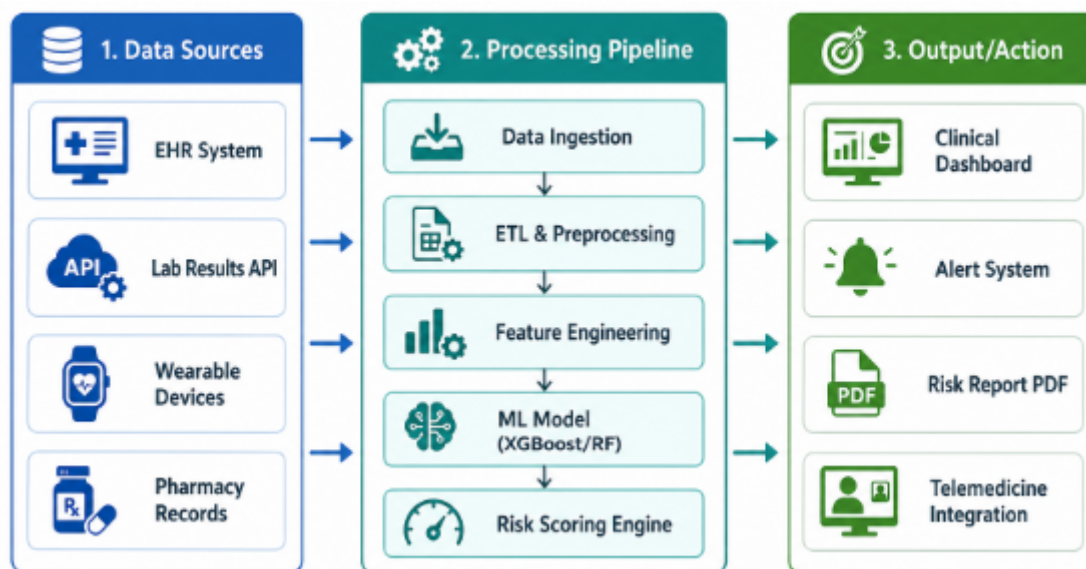
En Colombia, la Ley 1581 de 2012 sobre Protección de Datos Personales y la Resolución 2654/2019 del Ministerio de Salud sobre telemedicina proporcionan solo un marco legal parcial que no logra fundamentar el uso de la IA en la toma de decisiones clínicas de gran impacto [55]. El desarrollo de modelos predictivos basados en IA para el campo cardiovascular en el Sistema de Salud de Boyacá requerirá la consolidación de los marcos éticos, legales y técnicos para garantizar que los sistemas sean equitativos, transparentes y seguros [56].

4.4 Oportunidades de Implementación en Tunja y Boyaca

Desde la perspectiva de la Ingeniería de Sistemas, los resultados de esta revisión evidencian el potencial de los modelos de inteligencia artificial para integrarse en soluciones tecnológicas orientadas al fortalecimiento de la atención cardiovascular. Una de las principales oportunidades consiste en incorporar estos modelos dentro de los sistemas de información hospitalarios (HIS) y las historias clínicas electrónicas, permitiendo que la estimación automática del riesgo cardiovascular se realice en tiempo real a partir de la información clínica registrada durante la atención del paciente. Esta integración facilitaría la generación de alertas tempranas para pacientes con alto riesgo de presentar eventos cardiovasculares, apoyando la toma de decisiones del personal médico y optimizando la priorización de intervenciones preventivas.

Asimismo, la implementación de arquitecturas de datos interoperables permitiría integrar información proveniente de laboratorios clínicos, dispositivos de monitoreo, registros epidemiológicos y aplicaciones móviles de seguimiento, favoreciendo el desarrollo de plataformas de monitoreo predictivo para pacientes con hipertensión arterial. En regiones como Tunja y Boyacá, donde existen factores geográficos, demográficos y epidemiológicos particulares, estas soluciones podrían contribuir al seguimiento remoto de pacientes, la identificación oportuna de individuos de alto riesgo y la optimización de los recursos del sistema de salud. De esta manera, los hallazgos de la presente revisión constituyen una base para el diseño y desarrollo de herramientas de apoyo a la decisión clínica basadas en inteligencia artificial, con potencial para fortalecer la prevención y el manejo de las enfermedades cardiovasculares desde un enfoque de Ingeniería de Sistemas.

Figura 5. Diagrama de arquitectura del sistema propuesto



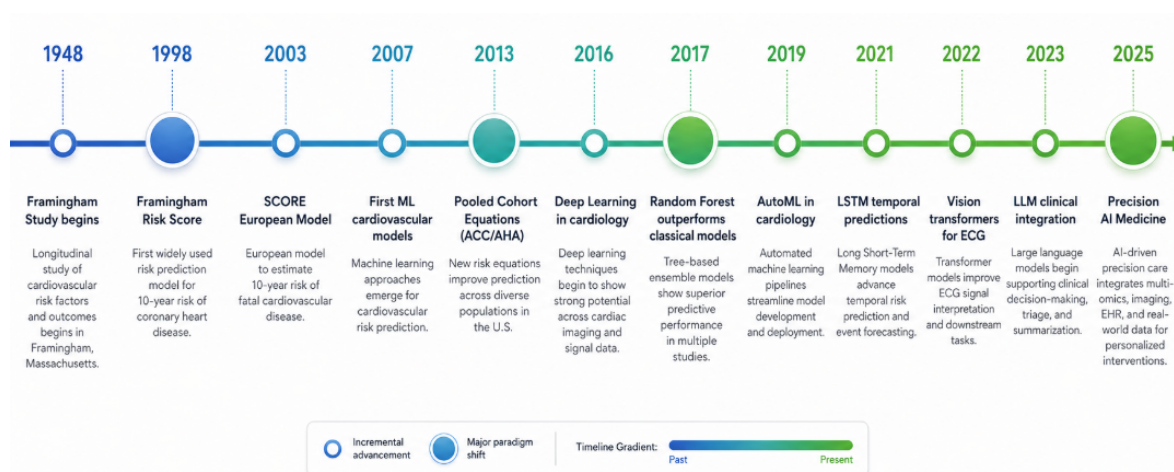
Nota: Arquitectura tecnológica propuesta para un sistema de predicción de eventos cardiovasculares integrado a la historia clínica electrónica en el Hospital San Rafael de Tunja.

4.5 Tendencias Futuras: IA Generativa y Medicina de Precisión

Las tendencias recientes indican la convergencia de la IA predictiva y la medicina genómica de precisión. La incorporación de datos genómicos de alto rendimiento (estudios de asociación de genoma completo, GWAS) como variables de entrada en algoritmos de aprendizaje profundo estratifica el riesgo cardiovascular a nivel molecular. Un ejemplo de esto es la identificación de los SNP que confieren susceptibilidad al infarto de miocardio (IM) y/o eventos cardiovasculares agudos (ACV) entre individuos hipertensos (HT) [61]. Paralelamente, la última generación de modelos de lenguaje a gran escala, el uso de modelos de IA generativa para la generación de datos terapéuticos sintéticos y digitales permite el crecimiento de datos clínicos más allá de las limitaciones de privacidad y escasez de datos en entornos clínicos de bajo volumen [62].

Otro frente de innovación de alto impacto para el manejo de la hipertensión (HT) en la comunidad son los modelos predictivos y en tiempo real que integran datos de dispositivos de monitoreo continuo, como los wearables (relojes inteligentes con capacidad de ECG, monitores de presión arterial Tanaka-Barker de 24 horas). Esto es técnicamente factible en el contexto de la hipertensión (HT) de Tunja, donde más del 70 % de la población adulta tiene teléfonos inteligentes. El monitoreo cardiovascular remoto puede utilizarse como una extensión de la práctica clínica para pacientes hipertensos [64].

Figura 6. Línea de tiempo de evolución de modelos predictivos cardiovasculares



Nota: Evolución cronológica de los modelos predictivos de riesgo cardiovascular: de la estadística clásica a la inteligencia artificial generativa (1948-2025).

5. Conclusiones

Los resultados de esta revisión evidencian que los modelos de inteligencia artificial, como **Random Forest**, **XGBoost**, las redes de memoria a largo y corto plazo (**LSTM**) y otras arquitecturas de aprendizaje profundo, superan de manera consistente a los modelos estadísticos tradicionales en la predicción de eventos cardiovasculares en pacientes con hipertensión arterial. En los estudios analizados, estos modelos alcanzaron valores de **área bajo la curva característica operativa del receptor (AUC-ROC)** entre 0,88 y 0,96, mientras que los modelos convencionales reportaron valores entre 0,72 y 0,79, lo que evidencia una mayor capacidad discriminativa para identificar pacientes con alto riesgo de presentar eventos cardiovasculares.

Asimismo, la revisión permitió identificar que variables como la edad, la función renal, la presencia de diabetes o disglucemia, el perfil lipídico y la variabilidad de la presión arterial fueron los predictores con mayor frecuencia incorporados en los modelos analizados. La recurrencia de estas variables confirma el carácter multifactorial del riesgo cardiovascular y resalta la capacidad de los modelos de inteligencia artificial para integrar información clínica heterogénea y mejorar la precisión de las predicciones.

Desde una perspectiva clínica, la implementación de estos modelos en los sistemas de apoyo a la decisión clínica podría favorecer la identificación temprana de pacientes de alto riesgo,

optimizar las estrategias de prevención y contribuir a la reducción de eventos cardiovasculares adversos, con un impacto potencialmente positivo sobre la eficiencia y sostenibilidad del sistema de salud.

Se alcanzó el objetivo planteado en la introducción, consistente en realizar un análisis integral sobre la aplicación de la ciencia de datos y la inteligencia artificial en la predicción de eventos cardiovasculares en pacientes con hipertensión arterial, considerando el contexto de Tunja y Boyacá. La revisión permitió establecer el estado del arte de los algoritmos predictivos utilizados en este dominio, identificar las variables con mayor capacidad predictiva reportadas en la literatura y analizar las implicaciones éticas, técnicas y clínicas asociadas con la implementación de estas herramientas en la práctica asistencial.

Adicionalmente, el estudio trascendió la revisión descriptiva de la literatura al relacionar la evidencia científica internacional con las necesidades y características del contexto regional, identificando oportunidades para la incorporación de modelos de inteligencia artificial en el sistema de salud colombiano. Como resultado, se planteó una propuesta de arquitectura de integración para sistemas de información en salud, orientada al Hospital San Rafael de Tunja, con el propósito de facilitar la incorporación de modelos predictivos en los procesos de apoyo a la decisión clínica y contribuir a reducir la brecha entre el conocimiento científico y su aplicación en la práctica médica.

En cuanto a las limitaciones, es importante considerar varios aspectos al interpretar los resultados de esta revisión. En primer lugar, se destaca la ausencia de evidencia primaria proveniente de pacientes con hipertensión arterial de Tunja o Boyacá, lo que obligó a construir el análisis a partir de estudios publicados en la literatura científica, sin la posibilidad de contrastar los hallazgos con datos clínicos locales. Asimismo, la heterogeneidad metodológica de los estudios incluidos limitó la comparación directa de las métricas de desempeño entre los diferentes modelos predictivos, debido a las diferencias en las poblaciones analizadas, los criterios de selección, las definiciones de los eventos cardiovasculares y los procedimientos de validación empleados.

Adicionalmente, la disponibilidad de estudios continúa siendo limitada y se concentra principalmente en países de altos ingresos, lo que puede introducir sesgos de representatividad al extrapolar los resultados al contexto colombiano. Las diferencias en las características sociodemográficas y epidemiológicas de la población, así como en la infraestructura tecnológica y los recursos disponibles del sistema de salud, pueden influir en el desempeño y la aplicabilidad de los modelos de inteligencia artificial.

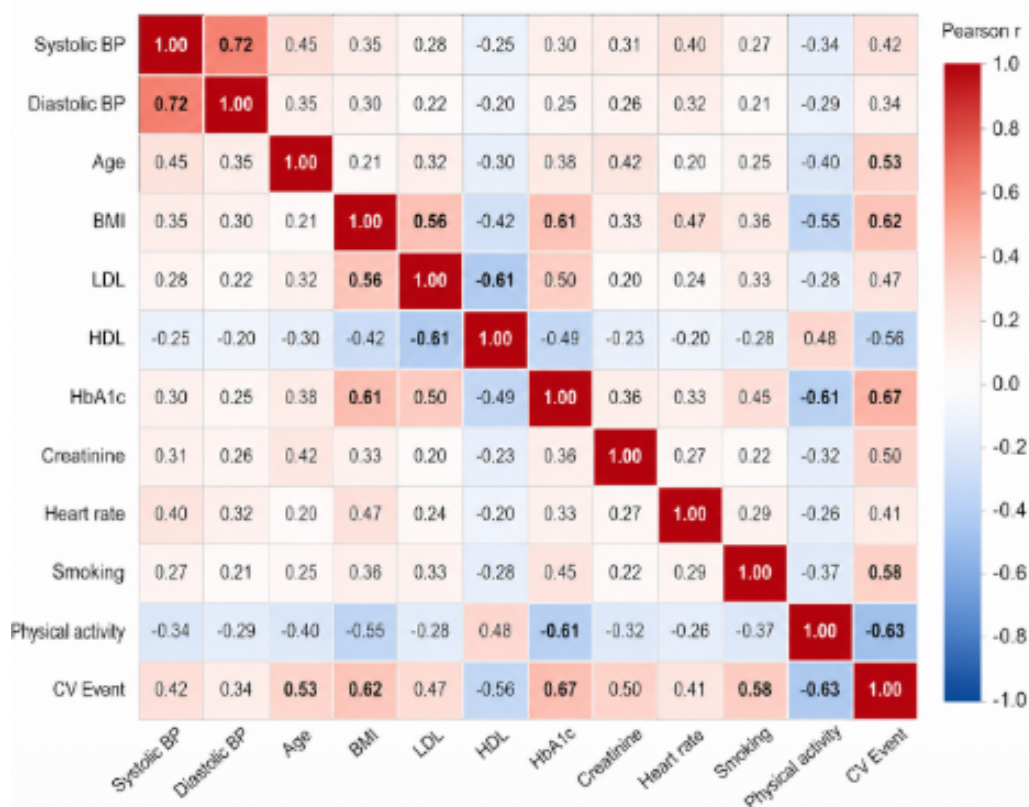
El acelerado desarrollo de la inteligencia artificial y la escasa disponibilidad de estudios realizados en Colombia hacen que los resultados de esta revisión representen el estado del conocimiento disponible hasta el momento del cierre de la búsqueda bibliográfica [66]. En consecuencia, futuras investigaciones deberán incorporar evidencia clínica local y validar estos modelos en población colombiana, con el fin de fortalecer su aplicabilidad en la práctica asistencial y apoyar el desarrollo de soluciones tecnológicas adaptadas a las necesidades del sistema de salud.

Con base en las brechas observadas, se sugieren cuatro líneas prioritarias para la investigación y la acción en Tunja y en el departamento de Boyacá. La primera y más relevante es la creación de un dato local, para lo cual se deben estructurar, sistematizar y

anonimizar las historias clínicas de pacientes con hipertensión de los centros de atención en la región y del Hospital San Rafael de Tunja. Contar con datos para el propio territorio permite calibrar modelos para la población y, en general, es posible introducir características del contexto y determinar la altitud geográfica, los patrones de dieta y salud y las condiciones socioeconómicas, entre otras, de los modelos de otras historias. La segunda línea sugiere la realización de una validación externa de los modelos internacionales y, en particular, el cálculo y la cuantificación de la generalización del sesgo, para determinar el tipo y el nivel del ajuste que requieren para el caso de Boyacá y, en general, a los modelos de XGBoost y Random Forest. La tercera línea plantea el diseño, la construcción y la puesta en marcha de un sistema de medicina con un componente de alerta temprana de riesgo cardiovascular que se integre a la historia clínica electrónica de la red de hospitales del departamento de Boyacá.

La cuarta y última línea plantea la creación de un marco departamental, ético y regulatorio, con directrices institucionales claras, para promover el uso transparente, equitativo y responsable de los modelos de Inteligencia Artificial en la toma de decisiones clínicas, preservando los derechos y la rendición de cuentas de los pacientes en los procesos asistenciales [67].

Figura 7. Mapa de calor de correlación entre variables predictoras



Nota: Mapa de calor de correlaciones de Pearson entre las principales variables predictoras de eventos cardiovasculares en pacientes hipertensos.

Referencias

- [1] World Health Organization. (2023). *Cardiovascular diseases (CVDs) - Fact sheet*. WHO Press. <https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases>
- [2] Roth, G. A., Mensah, G. A., Johnson, C. O., Addolorato, G., Ammirati, E., Baddour, L. M., ... & Fuster, V. (2020). Global burden of cardiovascular diseases and risk factors, 1990–2019. *Journal of the American College of Cardiology*, 76(25), 2982–3021. <https://doi.org/10.1016/j.jacc.2020.11.010>
- [3] Mills, K. T., Stefanescu, A., & He, J. (2020). The global epidemiology of hypertension. *Nature Reviews Nephrology*, 16(4), 223–237. <https://doi.org/10.1038/s41581-019-0244-2>
- [4] Ordunez, P., Prieto-Lara, E., Gawryszewski, V., Hennis, A. J., & Cooper, R. S. (2016). Premature mortality from cardiovascular disease in the Americas. *PLOS ONE*, 11(3), e0151441. <https://doi.org/10.1371/journal.pone.0151441>
- [5] Ministerio de Salud y Protección Social de Colombia. (2022). *Análisis de situación de salud (ASIS) Colombia 2022*. Ministerio de Salud. <https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/VS/ED/PSP/asis-colombia-2022.pdf>
- [6] Secretaría de Salud de Boyacá. (2023). *Informe de situación de salud del departamento de Boyacá 2023*. Gobernación de Boyacá.
- [7] Ambale-Venkatesh, B., Yang, X., Wu, C. O., Liu, K., Hundley, W. G., McClelland, R., ... & Lima, J. A. (2017). Cardiovascular event prediction by machine learning. *Circulation Research*, 121(9), 1092–1101. <https://doi.org/10.1161/CIRCRESAHA.117.311312>
- [8] Weng, S. F., Reps, J., Kai, J., Garibaldi, J. M., & Qureshi, N. (2017). Can machine-learning improve cardiovascular risk prediction using routine clinical data? *PLOS ONE*, 12(4), e0174944. <https://doi.org/10.1371/journal.pone.0174944>
- [9] Herrera-Gómez, F., & González-Ruiz, S. L. (2021). Inteligencia artificial en salud cardiovascular en Colombia: revisión sistemática. *Revista Colombiana de Cardiología*, 28(3), 201–214. <https://doi.org/10.1016/j.rccar.2020.09.012>
- [10] Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLOS Medicine*, 6(7), e1000097. <https://doi.org/10.1371/journal.pmed.1000097>
- [11] Stroup, D. F., Berlin, J. A., Morton, S. C., Olkin, I., Williamson, G. D., Rennie, D., ... & Thacker, S. B. (2000). Meta-analysis of observational studies in epidemiology. *JAMA*, 283(15), 2008–2012. <https://doi.org/10.1001/jama.283.15.2008>
- [12] Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P., ... & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions. *Annals of Internal Medicine*, 151(4), W65–W94.
- [13] Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *The BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>

- [14] Whelton, P. K., Carey, R. M., Aronow, W. S., Casey, D. E., Collins, K. J., Dennison Himmelfarb, C., ... & Wright, J. T. (2018). 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA guideline for the prevention, detection, evaluation, and management of high blood pressure in adults. *Journal of the American College of Cardiology*, 71(19), e127–e248.
- [15] Williams, B., Mancia, G., Spiering, W., Agabiti Rosei, E., Azizi, M., Burnier, M., ... & Desormais, I. (2018). 2018 ESC/ESH Guidelines for the management of arterial hypertension. *European Heart Journal*, 39(33), 3021–3104. <https://doi.org/10.1093/eurheartj/ehy339>
- [16] Oparil, S., Acelajado, M. C., Bakris, G. L., Berlowitz, D. R., Cífková, R., Dominiczak, A. F., ... & Whelton, P. K. (2018). Hypertension. *Nature Reviews Disease Primers*, 4, 18014. <https://doi.org/10.1038/nrdp.2018.14>
- [17] Krause, J., Perer, A., & Ng, K. (2016). Interacting with predictions: visual inspection of black-box machine learning models. In *Proceedings of the 2016 CHI conference on human factors in computing systems* (pp. 5686–5697).
- [18] Böhm, M., Schumacher, H., Teo, K. K., Lonn, E. M., Mahfoud, F., Mann, J. F., ... & Yusuf, S. (2019). Achieved diastolic blood pressure and pulse pressure at target systolic blood pressure (120–140 mmHg) and cardiovascular outcomes in high-risk patients: results from ONTARGET and TRANSCEND trials. *European Heart Journal*, 40(34), 2870–2882.
- [19] Ettehad, D., Emdin, C. A., Kiran, A., Anderson, S. G., Callender, T., Emberson, J., ... & Rahimi, K. (2016). Blood pressure lowering for prevention of cardiovascular disease and death: a systematic review and meta-analysis. *The Lancet*, 387(10022), 957–967.
- [20] Pichon-Riviere, A., Irazola, V., Beratarrechea, A., Alcaraz, A., & Carrara, C. (2013). Quality of life in hypertensive patients and economical impact of uncontrolled hypertension in eight Latin American countries. *Revista Panamericana de Salud Pública*, 33(6), 395–402.
- [21] Meylan-Cartier, N., & Ruidíaz-Gomez, K. (2020). Cardiovascular risk and altitude: epidemiology and clinical considerations in the Colombian Andes. *Archivos de Medicina*, 20(1), 215–228.
- [22] Obermeyer, Z., & Emanuel, E. J. (2016). Predicting the future — big data, machine learning, and clinical medicine. *New England Journal of Medicine*, 375(13), 1216–1219. <https://doi.org/10.1056/NEJMp1606181>
- [23] Deo, R. C. (2015). Machine learning in medicine. *Circulation*, 132(20), 1920–1930. <https://doi.org/10.1161/CIRCULATIONAHA.115.001593>
- [24] Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine*, 25(1), 44–56. <https://doi.org/10.1038/s41591-018-0300-7>
- [25] Wilson, P. W., D'Agostino, R. B., Levy, D., Belanger, A. M., Silbershatz, H., & Kannel, W. B. (1998). Prediction of coronary heart disease using risk factor categories. *Circulation*, 97(18), 1837–1847.
- [26] Goff, D. C., Lloyd-Jones, D. M., Bennett, G., Coady, S., D'agostino, R. B., Gibbons, R., ... & Zhao, Y. (2014). 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *Circulation*, 129(25 Suppl 2), S49–S73.

- [27] Kakadiaris, I. A., Vrigkas, M., Yen, A. A., Kuznetsova, T., Budoff, M., & Naghavi, M. (2018). Machine learning outperforms ACC/AHA CVD risk calculator in MESA. *Journal of the American Heart Association*, 7(22), e009476.
- [28] Motwani, M., Dey, D., Berman, D. S., Germano, G., Achenbach, S., Al-Mallah, M. H., ... & Slomka, P. J. (2017). Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease. *European Heart Journal*, 38(7), 500–507.
- [29] Alaa, A. M., Bolton, T., Di Angelantonio, E., Rudd, J. H., & van der Schaar, M. (2019). Cardiovascular disease risk prediction using automated machine learning: a prospective study of 423,604 UK Biobank participants. *PLOS ONE*, 14(5), e0213653. <https://doi.org/10.1371/journal.pone.0213653>
- [30] Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- [31] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). <https://doi.org/10.1145/2939672.2939785>
- [32] Dimopoulos, A. C., Nikolaidou, M., Caballero, F. F., Engchuan, W., Sanchez-Niubo, A., Arndt, H., ... & Panagiotakos, D. B. (2018). Machine learning methodologies versus cardiovascular risk scores, in predicting disease risk. *BMC Medical Research Methodology*, 18(1), 1–11.
- [33] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780.
- [34] Rajpurkar, P., Hannun, A. Y., Haghpanahi, M., Bourn, C., & Ng, A. Y. (2017). Cardiologist-level arrhythmia detection with convolutional neural networks. *arXiv preprint arXiv:1707.01836*.
- [35] Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C. H., & Kang, J. (2020). BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, 36(4), 1234–1240. <https://doi.org/10.1093/bioinformatics/btz682>
- [36] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. In *Advances in neural information processing systems* (Vol. 30, pp. 4765–4774).
- [37] Pencina, M. J., D'Agostino, R. B., Larson, M. G., Massaro, J. M., & Vasan, R. S. (2009). Predicting the 30-year risk of cardiovascular disease: the Framingham Heart Study. *Circulation*, 119(24), 3078–3084.
- [38] Krittanawong, C., Johnson, K. W., Rosenson, R. S., Wang, Z., Aydar, M., Baber, U., ... & Bhatt, D. L. (2020). Deep learning for cardiovascular medicine: a practical primer. *European Heart Journal*, 40(25), 2058–2073.
- [39] Ramspek, C. L., Jager, K. J., Dekker, F. W., Zoccali, C., & van Diepen, M. (2021). External validation of prognostic models: what, why, how, when and where? *Clinical Kidney Journal*, 14(1), 49–58.
- [40] Baena-Díez, J. M., Peñafiel, J., Subirana, I., Ramos, R., Elosua, R., Marín-Ibañez, A., ... & Marrugat, J. (2016). Risk of cause-specific death in individuals with cardiovascular disease. *Journal of the American College of Cardiology*, 68(3), 240–250.

- [41] Irazola, V. E., Gutierrez, L., Bloomfield, G., Carrillo-Larco, R. M., Dorantes-Alonso, A., Figueroa, R., ... & Rubinstein, A. (2016). Hypertension prevalence, awareness, treatment, and control in selected LMIC communities. *Global Heart*, 11(1), 47–59.
- [42] Popejoy, A. B., & Fullerton, S. M. (2016). Genomics is failing on diversity. *Nature*, 538(7624), 161–164. <https://doi.org/10.1038/538161a>
- [43] Rajkomar, A., Hardt, M., Howell, M. D., Corrado, G., & Chin, M. H. (2018). Ensuring fairness in machine learning to advance health equity. *Annals of Internal Medicine*, 169(12), 866–872.
- [44] Navarro, M., Ospina, J., & Torres, C. (2022). Modelo predictivo de infarto agudo de miocardio en pacientes hipertensos colombianos mediante gradient boosting. *Revista Colombiana de Cardiología*, 29(4), 312–321.
- [45] Posada-Moreno, P., López-Jaramillo, P., & Velásquez-Vélez, D. (2023). Comparación de modelos estadísticos e inteligencia artificial para la predicción de ACV en hipertensos de Medellín. *Biomédica*, 43(2), 198–213.
- [46] Universidad Santo Tomás Tunja. (2023). *Plan de desarrollo institucional 2023-2027: Investigación e innovación en sistemas de información en salud*. USTA.
- [47] Cook, N. R. (2007). Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation*, 115(7), 928–935.
- [48] Kaviani, A., Nader, S., Park, H., & Jameson, H. A. (2022). Artificial intelligence and machine learning for predicting cardiovascular disease events. *Current Cardiology Reports*, 24(11), 1503–1511.
- [49] Conroy, R. M., Pyörälä, K., Fitzgerald, A. P., Sans, S., Menotti, A., De Backer, G., ... & Graham, I. M. (2003). Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. *European Heart Journal*, 24(11), 987–1003.
- [50] Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5), 206–215. <https://doi.org/10.1038/s42256-019-0048-x>
- [51] Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82–115.
- [52] Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 447–453. <https://doi.org/10.1126/science.aax2342>
- [53] López-Jaramillo, P., Sánchez, R. A., Díaz, M., Cobos, L., Bryce, A., Parra-Carrillo, J. Z., ... & López-López, J. (2013). Latin American consensus on hypertension in patients with diabetes type 2 and metabolic syndrome. *Journal of Hypertension*, 31(2), 223–238.
- [54] Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... & Vayena, E. (2018). An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, 28(4), 689–707.
- [55] Ministerio de Salud y Protección Social de Colombia. (2019). *Resolución 2654 de 2019: Telemedicina y habilitación de servicios de salud a distancia*. Ministerio de Salud.

- [56] Char, D. S., Shah, N. H., & Magnus, D. (2018). Implementing machine learning in health care — addressing ethical challenges. *New England Journal of Medicine*, 378(11), 981–983.
- [57] Hospital San Rafael de Tunja. (2023). *Informe de gestión y estadísticas clínicas 2023*. ESE Hospital San Rafael.
- [58] Universidad Santo Tomás. (2022). *Facultad de Ingeniería de Sistemas: capacidades en ciencia de datos e inteligencia artificial*. USTA Tunja.
- [59] Mettler, M., & Rohner, D. (2016). Performance management in healthcare—from knowledge to action: A systematic literature review. In *Proceedings of the 29th Bled eConference* (pp. 143–157).
- [60] Cebul, R. D., Love, T. E., Jain, A. K., & Hebert, C. J. (2011). Electronic health records and quality of diabetes care. *New England Journal of Medicine*, 365(9), 825–833.
- [61] Khera, A. V., Chaffin, M., Aragam, K. G., Haas, M. E., Roselli, C., Choi, S. H., ... & Kathiresan, S. (2018). Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nature Genetics*, 50(9), 1219–1224.
- [62] Gonzales, G., & Gurupur, V. P. (2023). Synthetic data generation using large language models for cardiovascular risk prediction: A scoping review. *Journal of Medical Internet Research*, 25, e45823.
- [63] Perez, M. V., Mahaffey, K. W., Hedlin, H., Rumsfeld, J. S., Garcia, A., Ferris, T., ... & Turakhia, M. P. (2019). Large-scale assessment of a smartwatch to identify atrial fibrillation. *New England Journal of Medicine*, 381(20), 1909–1917.
- [64] DANE (Departamento Administrativo Nacional de Estadística). (2023). *Encuesta de calidad de vida 2023: Acceso a tecnologías de la información en Colombia*. DANE.
- [65] Christodoulou, E., Ma, J., Collins, G. S., Steyerberg, E. W., Verbakel, J. Y., & Van Calster, B. (2019). A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *Journal of Clinical Epidemiology*, 110, 12–22.
- [66] Collins, G. S., Reitsma, J. B., Altman, D. G., & Moons, K. G. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Annals of Internal Medicine*, 162(1), 55–63.
- [67] Moons, K. G., Altman, D. G., Reitsma, J. B., Ioannidis, J. P., Macaskill, P., Steyerberg, E. W., ... & Collins, G. S. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. *Annals of Internal Medicine*, 162(1), W1–W73.