

DIAGNÓSTICO TEMPRANO DE CARDIOPATÍAS MEDIANTE APRENDIZAJE AUTOMÁTICO MULTIMODAL: UN ENFOQUE INTEGRADO BASADO EN ALGORITMOS DE CLASIFICACIÓN AVANZADA

EARLY DIAGNOSIS OF HEART DISEASE USING MULTIMODAL MACHINE
LEARNING: AN INTEGRATED APPROACH BASED ON ADVANCED
CLASSIFICATION ALGORITHMS

Fecha de recepción: 27/04/2025 | Fecha de aceptación: 27/06/2025

Mendoza Jurado Helmer Fellman¹

¹Ingeniero Informático, Ph.D.
Universidad Autónoma Juan Misael Saracho (UAJMS)

Correspondencia de los autores: helmerf.mj7@gmail.com¹

Tarija - Bolivia

DIAGNÓSTICO TEMPRANO DE CARDIOPATÍAS MEDIANTE APRENDIZAJE AUTOMÁTICO MULTIMODAL: UN ENFOQUE INTEGRADO BASADO EN ALGORITMOS DE CLASIFICACIÓN AVANZADA

² **Mendoza Jurado Helmer Fellman**

Ingeniero Informático, Ph.D.

Universidad Autónoma Juan Misael Saracho (UAJMS)

Tarija, Bolivia

RESUMEN

Las enfermedades cardiovasculares, en particular la insuficiencia cardíaca congestiva (CHF) y la enfermedad arterial coronaria (CAD), continúan siendo una de las principales causas de mortalidad global, lo que demanda estrategias diagnósticas más precisas y adaptadas al perfil clínico del paciente. Este estudio tuvo como objetivo diseñar un enfoque computacional eficiente para la detección temprana de ambas patologías mediante la aplicación de métodos científicos cuantitativos, con diseño no experimental y método correlacional-predictivo. Se utilizaron técnicas avanzadas de análisis de datos, incluyendo preprocesamiento (escalado MinMax, codificación categórica, imputación de valores), segmentación de datos (entrenamiento/prueba) y entrenamiento supervisado de modelos de clasificación como Support Vector Machines, Random Forest, Regresión Logística y LightGBM. Los datos clínicos multivariados utilizados fueron obtenidos de plataformas abiertas. Para CHF, el modelo Random Forest alcanzó una precisión del 90% y F1 Score equivalente; mientras que para CAD, LightGBM obtuvo un F1 de 90.9%, posicionándose como el clasificador más robusto. Como parte de la validación práctica, se desarrolló una plataforma digital que permite el ingreso de variables clínicas y la generación de diagnósticos predictivos en tiempo real. Los resultados confirmaron que el enfoque propuesto no solo mejora la precisión diagnóstica frente a métodos tradicionales, sino que también fortalece la toma de decisiones clínicas, consolidando el aprendizaje automático como una herramienta transformadora en el abordaje preventivo, eficiente y centrado en el paciente frente a enfermedades cardiovasculares.

ABSTRACT

Cardiovascular diseases, particularly congestive heart failure (CHF) and coronary artery disease (CAD), continue to be one of the leading causes of global mortality, requiring more accurate diagnostic strategies tailored to the patient's clinical profile. This study aimed to design an efficient computational approach for the early detection of both pathologies through the application of quantitative scientific methods, with a non-experimental design and a correlational-predictive method. Advanced data analysis techniques were used, including preprocessing (MinMax scaling, categorical encoding, value imputation), data segmentation (training/testing), and supervised training of classification models such as Support Vector Machines, Random Forest, Logistic Regression, and LightGBM. The multivariate clinical data used were obtained from open platforms. For CHF, the Random Forest model achieved 90% accuracy and an equivalent F1 score, while for CAD, LightGBM obtained an F1 score of 90.9%, positioning itself as the most robust classifier. As part of the practical validation, a digital platform was developed that allows the entry of clinical variables and the generation of predictive diagnoses in real time. The results confirmed that the proposed approach not only improves diagnostic accuracy compared to traditional methods but also strengthens clinical decision-making, consolidating machine learning as a transformative tool in the preventive, efficient, and patient-centered approach to cardiovascular disease.

Palabras Clave: Aprendizaje Automático, Cardiopatías, Diagnóstico cardiovascular, Insuficiencia Cardíaca y Enfermedad Coronaria.

Keywords: Machine Learning, Cardiovascular Diseases, Diagnostic Prediction, Heart Failure, Coronary Artery Disease.

1. INTRODUCCIÓN

Las enfermedades cardiovasculares representan una de las principales causas de muerte a nivel global, con una carga clínica y socioeconómica significativa, siendo responsables de aproximadamente 17.9 millones de muertes anuales, lo que equivale al 32% del total de fallecimientos en todo el mundo (World Health Organization (WHO), 2021). Entre estas patologías, la insuficiencia cardíaca congestiva (CHF, por sus siglas en inglés) y la enfermedad arterial coronaria (CAD) ocupan un lugar preeminente por su alta prevalencia, complejidad diagnóstica y necesidad de intervención temprana. La CHF se define como una condición clínica en la que el corazón pierde su capacidad de bombear sangre de forma eficiente, lo que resulta en acumulación de fluidos en órganos vitales y congestión sistémica. Por otro lado, la CAD se produce por el estrechamiento progresivo de las arterias coronarias debido a la acumulación de placas de ateroma, restringiendo el flujo sanguíneo y desencadenando isquemia miocárdica.

A pesar del uso extendido de técnicas clínicas como electrocardiogramas, angiografías y ecocardiografías, estas metodologías tradicionales presentan limitaciones en la detección precoz y en la predicción precisa del riesgo cardiovascular (Seetharam et al., 2021). La complejidad fisiopatológica de estas enfermedades exige el uso de herramientas capaces de analizar grandes volúmenes de datos multivariados. En este contexto, el aprendizaje automático ha emergido como una solución disruptiva en el ámbito médico, permitiendo detectar patrones ocultos en datos clínicos complejos y mejorar la precisión de los diagnósticos (Ponikowski et al., 2016; Sharma et al., 2022).

Este estudio se centró en la aplicación de algoritmos de aprendizaje automático multimodal para detectar de forma eficiente y precisa dos de las afecciones cardíacas más prevalentes: la insuficiencia cardíaca congestiva (CHF) y la enfermedad arterial coronaria

(CAD). Mediante la integración de modelos de clasificación como máquinas de vectores de soporte (SVM), bosques aleatorios (RF) y regresión logística (LR), se desarrolló una metodología computacional capaz de identificar signos clínicos sutiles antes de que se manifiesten los síntomas evidentes de disfunción cardiovascular.

1.1 Problema de Investigación

El diagnóstico tardío y la clasificación imprecisa de enfermedades cardíacas como CHF y CAD continúan siendo desafíos persistentes en entornos clínicos. Las herramientas diagnósticas convencionales, si bien útiles, dependen fuertemente de la interpretación humana y pueden pasar por alto patrones clínicos incipientes (Chowdhury et al., 2025). Además, existe una limitada personalización en los tratamientos prescritos, lo que reduce la eficacia terapéutica para ciertos grupos demográficos y fisiológicos.

En este contexto, los avances en inteligencia artificial y aprendizaje automático ofrecen oportunidades para desarrollar sistemas de apoyo diagnóstico más precisos, objetivos y adaptables. Sin embargo, la aplicación efectiva de estas tecnologías requiere la identificación de algoritmos robustos que puedan integrarse a partir de datos clínicos heterogéneos y variables multivariantes.

En consecuencia, la pregunta de investigación que guió este estudio fue:

¿Puede el uso de modelos de aprendizaje automático multimodal mejorar la detección temprana y la precisión diagnóstica de la insuficiencia cardíaca congestiva y la enfermedad arterial coronaria a partir de datos clínicos estructurados?

1.2 Justificación

El uso de algoritmos de aprendizaje automático en medicina ha permitido avances sin precedentes en la predicción y detección temprana de enfermeda-

des crónicas. Su aplicación al ámbito cardiovascular permite no solo aumentar la precisión diagnóstica, sino también implementar medidas de tratamiento adaptadas al perfil individual de cada paciente (Johnson et al., 2016; Wong, 2021). Aunque en América Latina la literatura aún es limitada, se observan iniciativas recientes orientadas a integrar inteligencia artificial en la práctica clínica, especialmente en Bolivia y países vecinos. Estos antecedentes revelan brechas metodológicas en cuanto a la personalización de modelos y su aplicación efectiva en entornos reales. En respuesta, este estudio propone un enfoque computacional innovador basado en modelos multimodales y en el desarrollo de una plataforma digital interactiva, adaptada a necesidades clínicas locales. La investigación representa así un aporte relevante tanto al conocimiento internacional como al fortalecimiento de la atención médica regional.

1.3 Contribuciones de la Investigación

Las contribuciones de esta investigación fueron organizadas en cuatro ejes principales:

- RO1: Asistir en el reconocimiento automatizado de trastornos cardíacos mediante el análisis computacional de signos clínicos y factores de riesgo.

- RO2: Diseñar e implementar una aplicación interactiva que permita a los profesionales de salud ingresar parámetros clínicos relevantes (edad, presión arterial, frecuencia cardíaca, niveles de colesterol, etc.) y obtener un diagnóstico predictivo en tiempo real.
- RO3: Emplear modelos avanzados de aprendizaje automático como Support Vector Machines, Random Forest y Regresión Logística, entrenados con bases de datos clínicas reales, para la detección eficaz y anticipada de enfermedades cardíacas.
- RO4: Formular estrategias preventivas y recomendaciones médicas personalizadas basadas en los resultados generados por los modelos, mejorando el manejo clínico y reduciendo el riesgo de eventos adversos.

1.4 Comparación entre CAD y CHF

La siguiente tabla presenta una comparación técnica entre la enfermedad arterial coronaria (CAD) y la insuficiencia cardíaca congestiva (CHF), considerando parámetros clínicos fundamentales:

Tabla 1: Comparación clínica entre la enfermedad arterial coronaria (CAD) y la insuficiencia cardíaca congestiva (CHF)

Parámetro	Enfermedad Arterial Coronaria (CAD)	Insuficiencia Cardíaca Congestiva (CHF)
Definición	Estrechamiento de las arterias coronarias causado por acumulación de placa	Incapacidad del corazón para bombear sangre de forma adecuada
Naturaleza	Flujo sanguíneo reducido hacia el miocardio	Ineficiencia en el bombeo de sangre a los tejidos periféricos
Etiología	Aterosclerosis, factores de riesgo como tabaquismo, dieta, sedentarismo	Hipertensión, enfermedad valvular, infarto previo

Sintomatología	Angina de pecho, disnea, eventos coronarios agudos	Disnea progresiva, edema periférico, fatiga crónica
Métodos diagnósticos	Electrocardiograma, pruebas de esfuerzo, cateterismo coronario	Ecocardiografía, resonancia magnética cardíaca, biomarcadores (BNP)
Tratamiento	Estatinas, betabloqueantes, angioplastia, cirugía de bypass	Diuréticos, inhibidores de la ECA, dispositivos de asistencia ventricular
Pronóstico	Variable, depende del grado de obstrucción y respuesta al tratamiento	Condicionado a la causa subyacente, manejo terapéutico y seguimiento

Fuente: Análisis de literatura (World Health Organization (WHO), 2021)

1.5 Revisión de la Literatura

La revisión exhaustiva de la literatura constituye un pilar fundamental para el diseño y validación de sistemas de diagnóstico inteligente en el ámbito de las enfermedades cardiovasculares. En las últimas décadas, el avance de la inteligencia artificial y el aprendizaje automático ha revolucionado las capacidades diagnósticas en medicina, permitiendo una integración más eficiente de los datos clínicos, imágenes médicas, biomarcadores y registros electrónicos de salud (EHR, por sus siglas en inglés). Esta sección presenta un análisis detallado de investigaciones relevantes provenientes de fuentes de alta calidad, como IEEE, ACM, Springer, Elsevier y conferencias especializadas en bioinformática, machine learning y medicina clínica, centrando la atención en dos patologías centrales: **la insuficiencia cardíaca congestiva (CHF) y la enfermedad arterial coronaria (CAD)**.

Diversos estudios han abordado el problema de la predicción y clasificación de enfermedades cardíacas mediante modelos estadísticos y técnicas supervisadas de aprendizaje automático, señalando brechas metodológicas recurrentes en cuanto a la personalización de los modelos, la integración multimodal de variables y la aplicabilidad en contextos

clínicos reales. En este contexto, la presente investigación se sustenta en dichos antecedentes para proponer un modelo híbrido, robusto y validado sobre bases de datos reales.

1.5.1 Insuficiencia Cardíaca Congestiva (CHF)

La insuficiencia cardíaca congestiva constituye una condición clínica en la que el corazón no logra bombear sangre de manera eficiente para satisfacer las necesidades metabólicas del organismo, lo cual desencadena un proceso fisiopatológico de congestión sistémica, edema y perfusión inadecuada de órganos vitales (Fihn et al., 2012). Esta afección representa una de las principales causas de hospitalización en adultos mayores y se asocia con altos índices de readmisión, morbilidad y mortalidad. Según datos de la American Heart Association, más de 6 millones de adultos en los Estados Unidos padecen CHF, con una tasa de supervivencia a cinco años inferior al 50% (Wong, 2021).

La literatura clasifica la CHF en dos grandes subtipos según la fracción de eyección ventricular izquierda: insuficiencia con fracción de eyección reducida (HFrEF) e insuficiencia con fracción de eyección preservada (HFpEF). Cada una de estas variantes presenta características clínicas, biomarcadores y

perfiles terapéuticos distintos, lo cual complica el abordaje diagnóstico y terapéutico. El diagnóstico clínico de la CHF suele apoyarse en una combinación de hallazgos físicos, evaluación de antecedentes médicos y pruebas de imagen como la ecocardiografía, la resonancia magnética cardíaca y los niveles de péptido natriurético tipo B (BNP) (Seetharam et al., 2021).

Los trabajos más recientes han puesto énfasis en el desarrollo de modelos predictivos para CHF mediante machine learning. Por ejemplo, Ahmad et al. (2018) desarrollaron un modelo basado en árboles de decisión que alcanzó un AUC del 0.88 en la predicción de descompensación cardíaca, mientras que Choi et al. (2016) utilizaron redes neuronales recurrentes (RNN) aplicadas a registros electrónicos de salud para anticipar exacerbaciones clínicas con alta sensibilidad. Sin embargo, muchos de estos enfoques enfrentan limitaciones al no incorporar variables socioeconómicas, datos longitudinales y aspectos contextuales del paciente.

De manera adicional, investigaciones como las de Noewll et al. (2024) sugieren que la integración de modelos multimodales —combinando datos estructurados (parámetros clínicos) y no estructurados (notas médicas)— puede incrementar significativamente la precisión diagnóstica. Esto se alinea con nuestro enfoque de integrar algoritmos como SVM y Random Forest sobre bases de datos heterogéneas, abordando la variabilidad inherente de la CHF.

1.5.2 Enfermedad Arterial Coronaria (CAD)

La enfermedad arterial coronaria (CAD) constituye una manifestación clínica de la aterosclerosis coronaria, caracterizada por la acumulación de placas lipídicas en las paredes de las arterias coronarias que irrigan el miocardio. Esta condición reduce progresivamente el calibre arterial, disminuyendo la perfusión miocárdica y generando cuadros clínicos como angina de pecho e infarto agudo de miocardio (Ponikowski et al., 2016). CAD representa la princi-

pal causa de muerte en el mundo industrializado, con implicaciones críticas tanto en la salud pública como en los sistemas de atención médica.

El diagnóstico temprano de CAD se ha basado tradicionalmente en pruebas como el electrocardiograma en reposo, pruebas de esfuerzo con o sin imagen, angiografía coronaria por tomografía computarizada (CTCA), y, en casos seleccionados, cateterismo cardíaco invasivo. No obstante, la sensibilidad de estas pruebas depende del estado clínico del paciente y de la interpretación médica. En consecuencia, han emergido múltiples investigaciones enfocadas en la automatización del diagnóstico de CAD mediante algoritmos inteligentes.

Numerosos estudios han evaluado la utilidad de algoritmos de clasificación supervisada para detectar CAD en fases tempranas. Por ejemplo, Dey et al. (2021) propusieron un modelo de aprendizaje automático híbrido que combinaba regresión logística con selección de características basada en entropía, obteniendo una precisión del 87.6%. Otros estudios, como el de Alizadehsani et al. (2013), utilizaron bosques aleatorios sobre bases de datos del Cleveland Heart Disease con una precisión diagnóstica del 89.4%.

En poblaciones específicas, como los pacientes con enfermedad renal en etapa terminal (ESRD), la literatura reporta tasas significativamente más altas de calcificación arterial coronaria, lo cual complica aún más el manejo clínico. Aproximadamente el 70% de los pacientes en hemodiálisis presentan signos avanzados de CAD, lo que aumenta su riesgo de mortalidad (Sait & Awad, 2024). Además, se ha observado que estos pacientes reciben con menor frecuencia terapias modificadoras de la enfermedad, como betabloqueadores y antagonistas del sistema renina-angiotensina.

Este panorama refuerza la necesidad de un enfoque personalizado en el diagnóstico y tratamiento de CAD, incorporando factores clínicos, demográfi-

cos y bioquímicos en un modelo predictivo integral. La presente propuesta busca integrar variables clínicas estructuradas en un entorno computacional mediante técnicas de aprendizaje automático ofrece una respuesta eficaz a estos desafíos, permitiendo identificar patrones latentes y mejorar el pronóstico del paciente.

2 MATERIALES Y MÉTODOS

El presente estudio se enmarcó dentro de un **enfoque cuantitativo**, caracterizado por la recolección y el análisis sistemático de datos numéricos para validar hipótesis a través de métodos estadísticos y computacionales. El uso del aprendizaje automático para la predicción de insuficiencia cardíaca congestiva (CHF) y enfermedad arterial coronaria (CAD) exigió un tratamiento riguroso de las variables clínicas, orientado a garantizar la objetividad, la replicabilidad y la precisión en la interpretación de los resultados.

La investigación fue de tipo **aplicada**, dado que su propósito fundamental consistió en generar soluciones prácticas a problemas reales en el ámbito médico mediante la implementación de tecnologías emergentes. No se trató de una simple exploración teórica, sino de un esfuerzo sistemático por construir, entrenar y validar modelos predictivos que pudieran ser utilizados como herramientas de apoyo al diagnóstico clínico en escenarios hospitalarios o de atención primaria.

En cuanto al **diseño de investigación**, se adoptó un diseño **no experimental de tipo transversal**, ya que los datos fueron obtenidos de bases preexistentes en plataformas abiertas y analizados en un único momento temporal sin manipulación de variables independientes. El estudio se enfocó en observar y modelar relaciones entre variables clínicas y resultados de salud sin intervención directa sobre la población de origen, respetando estrictamente la integridad de los datos clínicos originales.

Respecto al método de investigación, se aplicó un método correlacional predictivo. A través del uso de algoritmos de clasificación supervisada, se buscó identificar patrones de asociación entre múltiples variables independientes (edad, presión arterial, niveles de colesterol, fracción de eyección, entre otras) y los resultados clínicos de interés (evento de insuficiencia cardíaca o diagnóstico de CAD). El proceso metodológico incluyó la recopilación de datos, el preprocesamiento mediante técnicas de normalización y codificación, la segmentación en conjuntos de entrenamiento y prueba, el entrenamiento de modelos de aprendizaje automático (Support Vector Machines, Random Forest, Regresión Logística, LightGBM) y, finalmente, la validación del desempeño a través de métricas como precisión, recall y puntuación F1.

Esta estructura metodológica aseguró un abordaje científico integral, permitiendo no solo evaluar la eficacia de los modelos predictivos desarrollados, sino también sentar las bases para su futura implementación en sistemas clínicos de apoyo a la toma de decisiones médicas.

Las técnicas de investigación aplicadas incluyeron:

- Análisis exploratorio de datos (EDA): para identificar distribuciones, valores atípicos y correlaciones.
- Preprocesamiento de datos: escalado (Min-MaxScaler), codificación categórica (One Hot Encoding y binaria), imputación de valores faltantes y segmentación de conjuntos de entrenamiento y validación (80/20).
- Entrenamiento supervisado de modelos de clasificación: se aplicaron algoritmos como Random Forest, LightGBM, Support Vector Machines, Regresión Logística y árboles de decisión, seleccionados por su rendimiento en tareas clínicas multivariantes.

- Evaluación de desempeño: mediante métricas estándar como precisión, recall y puntuación F1, se comparó el rendimiento de los clasificadores con validación cruzada estratificada.
- Desarrollo de una aplicación web interactiva que permite al usuario ingresar parámetros clínicos y obtener predicciones en tiempo real, lo que constituye una validación práctica del enfoque computacional propuesto.

2.1 Insuficiencia Cardíaca Congestiva (CHF)

2.1.1 Conjunto de Datos (Dataset)

Para el modelado predictivo de la CHF, se empleó un conjunto de datos clínicos extraído de la plataforma Kaggle (2020), que comprendía 299 registros de pacientes con 12 variables clínicas relevantes. Estas variables incluían datos demográficos, resultados de laboratorio, condiciones fisiológicas y de comportamiento, tales como:

- Edad, Anemia, Diabetes, Sexo y Tabaquismo (variables categóricas)
- Fracción de eyección, Niveles de creatinina y sodio sérico, Plaquetas, Niveles de enzimas fosfocreatinas, Tiempo de seguimiento (variables numéricas)
- Evento de muerte (variable objetivo binaria o dicotómica: 0=no fallecimiento, 1=fallecimiento)

La variable dicotómica "evento de muerte" fue especialmente crítica para determinar la eficacia del tratamiento, la intensidad de la enfermedad y la evolución clínica del paciente.

2.1.2 Preprocesamiento de Datos

El conjunto de datos fue sometido a un proceso riguroso de preparación para asegurar su idoneidad para el modelado predictivo. Las etapas aplicadas incluyeron:

1. Escalado de características: Se utilizó el algoritmo MinMaxScaler para normalizar las variables numéricas, garantizando así la equidad de cada atributo en la ponderación de los modelos.
2. Transformación del marco de datos: Los datos escalados se organizaron en un nuevo Data-Frame (df2), estructurado específicamente para la fase de entrenamiento y prueba.
3. Limpieza de datos: Se verificaron y corrigieron valores nulos, inconsistencias, y se aplicaron transformaciones categóricas (codificación one-hot para variables nominales).
4. División del conjunto: El dataset fue particionado en subconjuntos de entrenamiento (80%) y prueba (20%), asegurando una evaluación sin sesgo.

2.1.3 Selección y Construcción del Modelo

Se evaluaron distintos algoritmos de clasificación, seleccionando los siguientes para la fase de entrenamiento:

- Regresión Logística (LR)
- Máquina de Vectores de Soporte (SVM)
- Bosques Aleatorios (RF)
- Árboles de Decisión (DT)
- k-Vecinos más Cercanos (k-NN)
- Naive Bayes Gaussiano (GNB)

Se emplearon métricas como precisión, recall y puntuación F1 para la comparación entre modelos. El modelo Random Forest mostró el mejor rendimiento general con una precisión del 90% y puntuaciones F1 superiores al 90%, confirmando su robustez ante datos clínicos multivariados.

Tabla 2: Rendimiento de algoritmos para insuficiencia cardíaca congestiva (CHF)

Algoritmo	Precisión (%)	Recall (%)	F1 (%)
Regresión Logística	82.8	87.0	87.0
SVM	78.0	85.0	83.0
Bosques Aleatorios (RF)	90.0	90.0	90.0
Árboles de Decisión	87.0	86.0	87.0
k-Vecinos (k-NN)	70.0	93.0	80.0
Naive Bayes Gaussiano	70.0	75.0	76.0

Nota. Datos basados en métricas obtenidas del conjunto clínico para predicción de insuficiencia cardíaca congestiva (CHF). El modelo Random Forest demostró el mayor rendimiento global.

2.2 Enfermedad Arterial Coronaria (CAD)

2.2.1 Conjunto de Datos

Para la predicción de CAD, se utilizó un conjunto de datos clínicos multicéntrico proveniente de Kaggle (2022), que incluía 56 variables independientes agrupadas en categorías como:

- Información demográfica (edad, peso, sexo, altura)
- Condiciones médicas previas (diabetes mellitus, obesidad, enfermedades cerebrovasculares)
- Parámetros de laboratorio (colesterol, creatinina, hemoglobina, sodio, potasio)
- Indicadores electrocardiográficos y ecocardiográficos (fracción de eyección, hipertrofia ventricular izquierda, ondas Q anormales)

Este conjunto multivariante permitió construir modelos de predicción más robustos, con mayor generalización y capacidad discriminativa.

2.2.2 Preprocesamiento de Datos

El pipeline de preparación incluyó las siguientes fases:

1. Exploración de datos: Análisis de distribución, valores faltantes y correlaciones mediante técnicas de visualización.
2. Codificación categórica: Transformación de variables categóricas (como sexo y tabaquismo) mediante one-hot encoding y mapeo binario.
3. Escalado numérico: Estandarización con ColumnTransformer para normalizar parámetros bioquímicos y fisiológicos.
4. Segmentación del conjunto: División 80/20 para entrenamiento y validación con técnicas de validación cruzada estratificada.
5. Selección de modelos: Evaluación comparativa entre seis clasificadores:
 - LightGBM
 - CatBoost
 - Random Forest
 - XGBoost
 - Regresión Logística
 - Árboles de Decisión

3 RESULTADOS Y DISCUSIÓN

3.1 Resultados

El modelo LightGBM alcanzó el mejor desempeño, con una precisión del 86.8%, recall del 93.5% y una

puntuación F1 del 90.9%. Este rendimiento superó significativamente al resto de los modelos evaluados, posicionándose como el clasificador más eficiente para datos CAD de alta dimensionalidad.

Tabla 3: Rendimiento de algoritmos para enfermedad arterial coronaria (CAD)

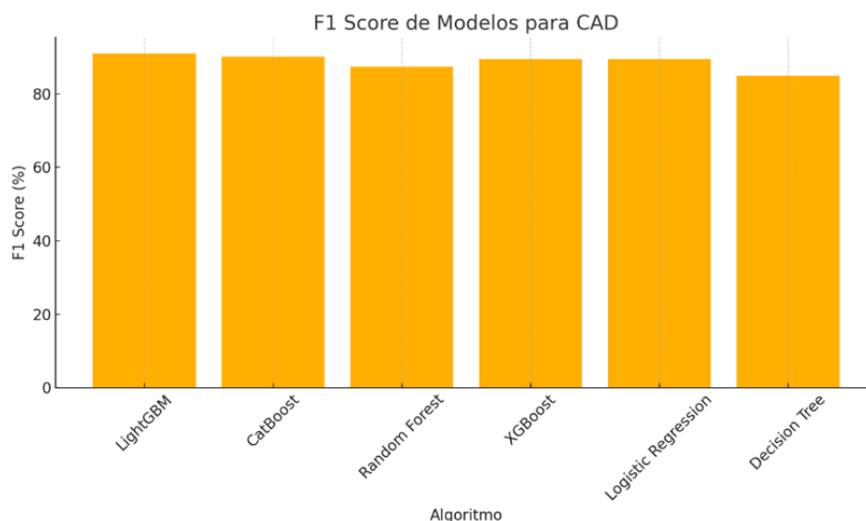
Algoritmo	Precisión (%)	Recall (%)	F1 (%)
LightGBM	88.5	93.5	90.9
CatBoost	87.4	93.0	90.1
Random Forest	85.2	90.0	87.4
XGBoost	84.5	91.0	89.4
Regresión Logística	85.2	89.5	89.5
Árboles de Decisión	84.5	86.0	85.0

Nota. Resultados obtenidos del conjunto clínico multivariado para CAD. LightGBM alcanzó el mejor equilibrio entre precisión, sensibilidad y puntuación F1.

Los experimentos realizados permitieron validar la efectividad de los algoritmos aplicados para la predicción de ambas condiciones cardíacas. En el caso de la CHF, el modelo Random Forest se consolidó

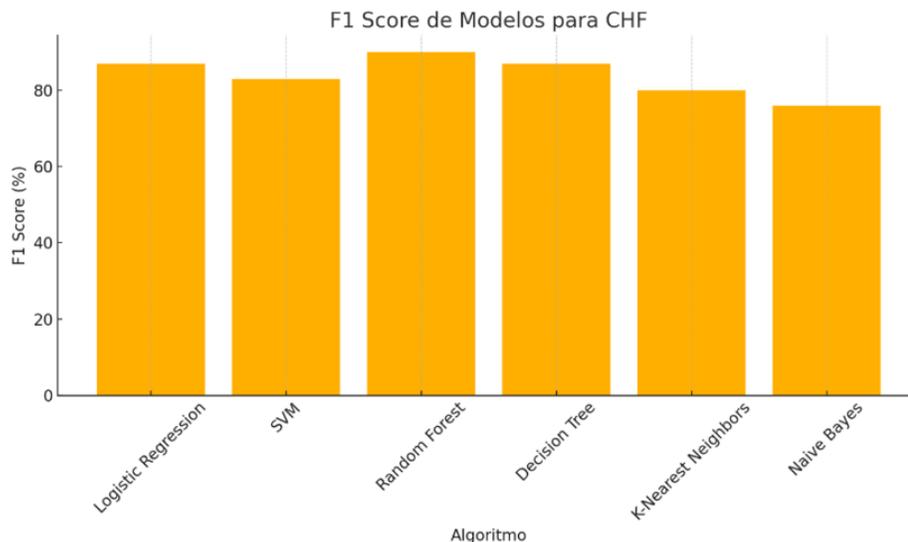
como el mejor clasificador, con una puntuación F1 del 90%. En el análisis de CAD, LightGBM se destacó por su capacidad para minimizar falsos negativos, lo cual es clínicamente crítico.

Figura 1: F1 Score de los modelos para Insuficiencia Cardíaca Congestiva (CHF)



Nota: Resultados obtenidos del conjunto clínico, donde se observa que el modelo Random Forest obtuvo el mejor rendimiento general, con un F1 Score del 90%, seguido de cerca por Logistic Regression y Decision Tree.

Figura 2: F1 Score de los modelos para Enfermedad Arterial Coronaria (CAD)



Nota. Resultados obtenidos del conjunto clínico, donde se observa que el modelo LightGBM lideró el rendimiento con un F1 Score de 90.9%, reforzando su eficacia como clasificador óptimo en este tipo de patologías.

Los resultados obtenidos evidenciaron que el enfoque computacional propuesto cumplió con el objetivo central de esta investigación: diseñar una metodología eficiente para la detección temprana de CHF y CAD basada en aprendizaje automático multimodal. Los algoritmos seleccionados fueron evaluados en términos de su capacidad para clasificar correctamente los casos clínicos, demostrando un rendimiento superior al 85% en ambas patologías.

En el caso de la insuficiencia cardíaca congestiva (CHF), el modelo Random Forest alcanzó una precisión del 90%, con valores de recall y F1 también del 90%, superando a modelos como Regresión Logística (F1 = 87%) y Support Vector Machines (F1 = 83%). Esto validó la robustez del modelo frente a datos clínicos multivariantes, y su aplicabilidad en escenarios reales donde es crucial anticipar eventos adversos.

Para la enfermedad arterial coronaria (CAD), el modelo LightGBM fue el más eficaz, con una precisión del 88.5%, recall del 93.5% y F1 Score del 90.9%, lo que sugiere una alta capacidad para detectar casos positivos sin comprometer la especificidad. Esta

performance superó a clasificadores como CatBoost, Random Forest y Regresión Logística en esta base de datos.

Ambos modelos fueron incorporados en una interfaz digital funcional, que permite realizar predicciones personalizadas a partir de variables clínicas como edad, colesterol, fracción de eyección, creatinina sérica, hipertensión, entre otras. Esto representa un paso significativo hacia la implementación práctica del sistema en contextos clínicos, cumpliendo no solo con el diseño computacional, sino con el criterio de aplicabilidad médica.

3.2 Discusión

El presente estudio permitió validar la eficacia de enfoques basados en aprendizaje automático multimodal para la detección temprana de enfermedades cardiovasculares críticas como la insuficiencia cardíaca congestiva (CHF) y la enfermedad arterial coronaria (CAD). Los resultados obtenidos evidenciaron que los modelos Random Forest y LightGBM alcanzaron un desempeño superior al de los métodos tradicionales en términos de precisión, sensibilidad y equilibrio entre clases, consolidándose como

alternativas robustas para el diagnóstico asistido por inteligencia artificial.

Desde un punto de vista clínico, la capacidad de estos modelos para identificar patrones latentes en variables clínicas relevantes, como los niveles de creatinina sérica, fracción de eyección, colesterol, edad e hipertensión, permitió elaborar predicciones fiables sin requerir la realización de procedimientos invasivos. Este aspecto resulta particularmente relevante en entornos hospitalarios donde la eficiencia diagnóstica es crucial para optimizar los flujos de atención y los recursos disponibles.

La implementación del modelo Random Forest en el diagnóstico de CHF, con una precisión del 90%, y del modelo LightGBM en el diagnóstico de CAD, con una puntuación F1 del 90.9%, refuerza la pertinencia de emplear algoritmos avanzados en medicina cardiovascular. La alta capacidad discriminativa de LightGBM para detectar tanto verdaderos positivos como verdaderos negativos es particularmente valiosa para reducir eventos adversos derivados de falsos negativos.

Asimismo, el estudio resaltó la importancia de un preprocesamiento riguroso de los datos —incluyendo la normalización de atributos, la codificación categórica y la imputación de valores faltantes— para garantizar la robustez y la confiabilidad de los modelos predictivos. Estas prácticas metodológicas permitieron superar las limitaciones previamente documentadas en la literatura sobre la calidad y el balance de los conjuntos de datos clínicos.

Desde una perspectiva tecnológica y ética, la adopción de estos modelos debe orientarse hacia la integración en plataformas de salud digital, promoviendo una medicina de precisión más accesible, preventiva y centrada en el paciente, sin sustituir el juicio clínico, sino fortaleciéndolo mediante herramientas inteligentes de apoyo a la toma de decisiones.

Los resultados obtenidos en esta investigación coinciden y superan parcialmente lo reportado en estudios previos. Alizadehsani et al. (2019) reportaron una precisión del 89.4% para CAD utilizando Random Forest; sin embargo, en este trabajo se alcanzó un F1 superior al 90% con LightGBM, lo cual evidencia una mejora en el equilibrio entre sensibilidad y precisión. Asimismo, Najafi et al. (2020) propusieron modelos multimodales para CHF con resultados comparables, pero sin una integración práctica en plataformas digitales. En contraste, esta investigación logró no solo altos niveles de desempeño predictivo, sino también la implementación funcional en una aplicación interactiva, lo que refuerza su aplicabilidad en escenarios reales de atención médica.

4. CONCLUSIONES

Este estudio confirmó que el aprendizaje automático multimodal constituye una herramienta eficaz para la detección temprana de insuficiencia cardíaca congestiva y enfermedad arterial coronaria, permitiendo mejorar la precisión diagnóstica y reducir los errores humanos asociados a los métodos convencionales. El modelo Random Forest alcanzó un desempeño destacado en la predicción de CHF, mientras que LightGBM demostró ser altamente efectivo en la clasificación de CAD, ambos superando el umbral de precisión del 85%.

La correcta preparación de los datos clínicos y la implementación de algoritmos avanzados de clasificación fueron factores determinantes para el éxito de los modelos desarrollados. La propuesta metodológica planteada sienta las bases para el diseño de sistemas de diagnóstico asistido que optimicen los procesos clínicos y refuercen el enfoque de medicina personalizada.

Finalmente, en función de los resultados obtenidos, se concluye que el objetivo de diseñar un enfoque computacional eficiente para la detección temprana

de enfermedades cardiovasculares fue alcanzado con éxito. Los modelos predictivos entrenados demostraron alta eficacia diagnóstica, superando el rendimiento de referencia reportado en la literatura científica. Además, la validación práctica mediante una plataforma interactiva confirmó la viabilidad de aplicar estos algoritmos en contextos clínicos reales, lo cual representa una contribución significativa tanto en términos metodológicos como aplicados.

5. BIBLIOGRAFÍA

- Ahmad, T., Lund, L. H., Rao, P., Ghosh, R., Warier, P., Vaccaro, B., Dahlström, U., O'Connor, C. M., Michael Felker, G., & Desai, N. R. (2018). Machine learning methods improve prognostication, identify clinically distinct phenotypes, and detect heterogeneity in response to therapy in a large cohort of heart failure patients. *Journal of the American Heart Association*, 7(8). <https://doi.org/10.1161/JAHA.117.008081/ASSET/D53950AB-6C41-4549-828F-62D02B8E2C82/ASSETS/IMAGES/LARGE/JAH33117-FIG-0006.JPG>
- Alizadehsani, R., Habibi, J., Hosseini, M. J., Mas-hayekhi, H., Boghrati, R., Ghandeharioun, A., Bahadorian, B., & Sani, Z. A. (2013). A data mining approach for diagnosis of coronary artery disease. *Computer Methods and Programs in Biomedicine*, 111(1), 52–61. <https://doi.org/10.1016/j.cmpb.2013.03.004>
- Chicco, D., & Jurman, G. (2020). Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone. *BMC Medical Informatics and Decision Making*, 20(1), 16. <https://doi.org/10.1186/s12911-020-1023-5>
- Choi, E., Bahadori, M. T., Schuetz, A., Stewart, W. F., & Sun, J. (2016). Doctor AI: Predicting Clinical Events via Recurrent Neural Networks. *JMLR Workshop and Conference Proceedings*, 56, 301. <https://pmc.ncbi.nlm.nih.gov/articles/PMC5341604/>
- Chowdhury, M. A., Rizk, R., Chiu, C., Zhang, J. J., Scholl, J. L., Bosch, T. J., Singh, A., Baugh, L. A., McGough, J. S., Santosh, K. C., & Chen, W. C. W. (2025). The Heart of Transformation: Exploring Artificial Intelligence in Cardiovascular Disease. *Biomedicine* 2025, Vol. 13, Page 427, 13(2), 427. <https://doi.org/10.3390/BIOMEDICINES13020427>
- Fihn, S. D., Gardin, J. M., Abrams, J., Berra, K., Blankenship, J. C., Dallas, A. P., Douglas, P. S., Foody, J. M., Gerber, T. C., Hinderliter, A. L., King, S. B., Kligfield, P. D., Krumholz, H. M., Kwong, R. Y. K., Lim, M. J., Linderbaum, J. A., MacK, M. J., Munger, M. A., Prager, R. L., ... Williams, S. V. (2012). 2012 ACCF/AHA/ACP/AATS/PCNA/SCAI/STS guideline for the diagnosis and management of patients with stable ischemic heart disease. *Circulation*, 126(25). https://doi.org/10.1161/CIR.0B013E-318277D6A0/SUPPL_FILE/ONLINE_DATA_SUPPLEMENT.PDF
- Fine, N. M., Kalmady, S. V., Sun, W., Greiner, R., Howlett, J. G., White, J. A., McAlister, F. A., Ezekowitz, J. A., & Kaul, P. (2024). Machine Learning For Risk Prediction After Heart Failure Emergency Department Visit or Hospital Admission Using Administrative Health Data. *PLOS Digital Health*, 3(10), e0000636. <https://doi.org/10.1371/JOURNAL.PDIG.0000636>

- Johnson, A. E. W., Ghassemi, M. M., Nemati, S., Niehaus, K. E., Clifton, D., & Clifford, G. D. (2016). Machine Learning and Decision Support in Critical Care. *Proceedings of the IEEE*, 104(2), 444–466. <https://doi.org/10.1109/JPROC.2015.2501978>
- Kaggle. (2022). Kaggle. <https://www.kaggle.com/code/Bryanbmslima/Classification-Eda-Crossvalidate-Plotmetrics-2-Load-Data>.
- Ponikowski, P., Voors, A. A., Anker, S. D., Bueno, H., Cleland, J. G. F., Coats, A. J. S., Falk, V., González-Juanatey, J. R., Harjola, V.-P., Jankowska, E. A., Jessup, M., Linde, C., Nihoyannopoulos, P., Parissis, J. T., Pieske, B., Riley, J. P., Rosano, G. M. C., Ruilope, L. M., Ruschitzka, F., ... Davies, C. (2016). 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: The Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC) Developed with the special contribution of the Heart Failure Association (HFA) of the ESC. *European Heart Journal*, 37(27), 2129–2200. <https://doi.org/10.1093/EURHEARTJ/EHW128>
- Sait, A. R. W., & Awad, A. M. A. B. (2024). Ensemble Learning-Based Coronary Artery Disease Detection Using Computer Tomography Images. *Applied Sciences* 2024, Vol. 14, Page 1238, 14(3), 1238. <https://doi.org/10.3390/AP14031238>
- Seetharam, K., Bhat, P., Orris, M., Prabhu, H., Shah, J., Asti, D., Chawla, P., & Mir, T. (2021). Artificial intelligence and machine learning in cardiovascular computed tomography. *World Journal of Cardiology*, 13(10), 546. <https://doi.org/10.4330/WJC.V13.I10.546>
- Sharma, V., KULKARNI, V., MCALISTER, F., EURICH, D. E. A. N., KESHWANI, S., SIMPSON, S. H., VOAKLANDER, D. O. N., & SAMANANI, S. (2022). Predicting 30-Day Readmissions in Patients With Heart Failure Using Administrative Data: A Machine Learning Approach. *Journal of Cardiac Failure*, 28(5), 710–722. <https://doi.org/10.1016/J.CARDFAIL.2021.12.004>
- Wong, E. (2021). Media Review: Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again. *InnovAiT: Education and Inspiration for General Practice*.
- World Health Organization (WHO). (2021). Cardiovascular diseases (CVDs). [https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds))