

<https://doi.org/10.69639/arandu.v11i2.401>

Análisis y Visualización de Datos Clínicos: Enfoque Analítico para la Predicción de Insuficiencia Cardíaca

Analysis and Visualization of Clinical Data: An Analytical Approach for Heart Failure Prediction

Erika Cristina Lozada Martínez

erikacristina.lozada@alumni.urv.cat

<https://orcid.org/0000-0001-8819-2366>

Universitat Rovira i Virgili

Tarragona – España

José Ezequiel Naranjo Robalino

jose.naranjo0463@utc.edu.ec

<https://orcid.org/0000-0002-2884-1667>

Universidad Técnica de Cotopaxi (UTC)

Ecuador – Latacunga

Marcelo Vladimir García Sánchez

[mv.garcia@uta.edu.ec](mailto:m.v.garcia@uta.edu.ec)

<https://orcid.org/0000-0002-7138-3913>

Universidad Técnica de Ambato

Ambato – Ecuador

Artículo recibido: 20 octubre 2024 -

Aceptado para publicación: 25 noviembre 2024

Conflictos de intereses: Ninguno que declarar

RESUMEN

Este estudio analiza el Heart Failure Prediction Dataset para identificar factores de riesgo asociados con la insuficiencia cardíaca mediante técnicas de visualización y análisis estadístico. El dataset incluye 11 características clínicas relevantes, como edad, género, tipo de dolor torácico, presión arterial en reposo, y otros indicadores cardiovasculares. Se preprocesaron los datos para garantizar su calidad, incluyendo la creación de variables lógicas y la segmentación por género, además de la eliminación de valores faltantes. Se emplearon diferentes tipos de visualizaciones, como histogramas, diagramas de caja y dispersión, para analizar la relación entre las variables y la probabilidad de insuficiencia cardíaca. Los resultados destacan que los hombres mayores de 55 años, con dolor torácico atípico (ASY) y angina inducida por ejercicio, presentan un riesgo significativamente mayor de desarrollar insuficiencia cardíaca. Otras variables, como el colesterol, la frecuencia cardíaca máxima y el nivel de glucosa en ayunas, no mostraron una influencia relevante en este contexto. Además, la pendiente del segmento ST (Slope) y la presión arterial en reposo en mujeres se identificaron como factores importantes en la predicción de la enfermedad. Este análisis refuerza la importancia de enfoques basados en datos para comprender mejor los factores de riesgo cardiovascular y desarrollar estrategias de prevención personalizadas.

Palabras clave: analítica de datos, diagnóstico clínico, enfermedad cardíaca

ABSTRACT

This study analyzes the Heart Failure Prediction Dataset to identify risk factors associated with heart failure using visualization techniques and statistical analysis. The dataset includes 11 relevant clinical characteristics, such as age, gender, type of chest pain, resting blood pressure, and other cardiovascular indicators. The data were preprocessed to ensure data quality, including the creation of logical variables and segmentation by gender, as well as the elimination of missing values. Different types of visualizations, such as histograms, box plots and scatter plots, were used to analyze the relationship between variables and the probability of heart failure. The results highlight that men older than 55 years, with atypical chest pain (ASY) and exercise-induced angina, have a significantly higher risk of developing heart failure. Other variables, such as cholesterol, maximum heart rate and fasting glucose level, did not show a relevant influence in this context. In addition, ST-segment slope (Slope) and resting blood pressure in women were identified as important factors in predicting disease. This analysis reinforces the importance of data-driven approaches to better understand cardiovascular risk factors and develop personalized prevention strategies.

Keywords: data analysis, clinical diagnosis, cardiac disease

INTRODUCCIÓN

La insuficiencia cardíaca es una condición crónica que afecta a millones de personas en todo el mundo, representando una de las principales causas de mortalidad y morbilidad. Su diagnóstico oportuno y la identificación precisa de los factores de riesgo asociados son esenciales para reducir las tasas de complicaciones y mejorar la calidad de vida de los pacientes (Li & Xia Dou, 2025). En este contexto, los avances en analítica de datos y la disponibilidad de conjuntos de datos clínicos estructurados ofrecen oportunidades para explorar patrones complejos y generar conocimiento útil para la predicción y prevención de eventos cardíacos adversos (Aghapanah et al., 2025).

El presente estudio utiliza el conjunto de datos que ha sido diseñado y ampliamente utilizado para explorar y predecir eventos de insuficiencia cardíaca, proporcionando 11 características clínicas que permiten analizar factores de riesgo clave en pacientes que incluyen información demográfica, fisiológica y metabólica, como edad, género, tipo de dolor torácico, presión arterial en reposo, colesterol, niveles de azúcar en ayunas, electrocardiogramas, frecuencia cardíaca máxima, angina inducida por ejercicio, descenso del segmento ST durante el esfuerzo en relación con el reposo y la pendiente del ejercicio pico. Estas variables permiten investigar cómo las características individuales y combinadas de los pacientes están asociadas con la probabilidad de insuficiencia cardíaca (Ijuin et al., 2025).

El objetivo principal de este trabajo es realizar un análisis exploratorio y visual del conjunto de datos para identificar patrones y relaciones significativas entre las variables clínicas y la insuficiencia cardíaca. Además, se busca responder preguntas específicas que relacionen las características clínicas con el riesgo de insuficiencia cardíaca, tales como la influencia del género, la edad o el tipo de dolor torácico en la probabilidad de sufrir este evento. La metodología aplicada incluye la selección de gráficos apropiados basados en principios de percepción visual y excelencia gráfica, garantizando que las visualizaciones sean claras, precisas y representen fielmente los datos.

El procesamiento de datos inicial consistió en convertir las variables a formatos lógicos y asignar valores predeterminados a las celdas vacías, con el fin de evitar sesgos y asegurar la consistencia del análisis. Posteriormente, se realizó un análisis descriptivo de cada variable para comprender su distribución y relación con la insuficiencia cardíaca. Las visualizaciones seleccionadas siguen principios como la posición en una escala común, el uso de gráficos tridimensionales en comparaciones multivariantes y la optimización de la relación tinta-datos, eliminando elementos innecesarios y garantizando la claridad del mensaje.

Este enfoque no solo permite responder preguntas específicas sobre las características clínicas y su relación con la insuficiencia cardíaca, sino que también demuestra cómo la analítica visual puede ser una herramienta poderosa en el análisis de datos médicos. Los hallazgos

obtenidos destacan la importancia de elegir técnicas gráficas adecuadas para maximizar la comprensión de los patrones clínicos y respaldar la toma de decisiones médicas basadas en datos. Además, este análisis visual sienta las bases para futuros estudios que integren modelos predictivos avanzados, con el objetivo de mejorar la identificación temprana de riesgos en pacientes con predisposición a eventos cardíacos.

MATERIALES Y MÉTODOS

Conjunto de datos

El análisis presentado en este estudio se basa en el conjunto de datos de predicción de insuficiencia cardíaca, un conjunto de datos clínicos disponible en la plataforma Kaggle (Garai et al., 2024), que contiene información de 303 pacientes, con un total de 12 columnas, de las cuales 11 corresponden a características clínicas y la restante a la variable objetivo (Cardiopatía), que indica la presencia (1) o ausencia (0) de insuficiencia cardíaca. Las características incluyen tanto variables numéricas como categóricas, relacionadas con factores de riesgo cardiovasculares ampliamente estudiados. El objetivo principal al procesar este conjunto de datos fue explorar las relaciones entre las variables para identificar patrones relevantes en la predicción de insuficiencia cardíaca. A continuación, se presenta la Tabla 1 con la descripción de las columnas del conjunto de datos.

Tabla 1

Descripción de conjunto de datos original

Variable	Dataset title	Tipo
Edad	Edad	Numérica
Sexo	Sexo	Categórica
Tipo de dolor en el pecho	TipoDolorPecho	Categórica
Presión arterial en reposo	PresionArterialReposo	Numérica
Colesterol sérico	Colesterol	Numérica
Azúcar en sangre en ayunas	AzucarSangreAyunas	Categórica
Resultados del electrocardiograma en reposo	ECGReposo	Categórica
Frecuencia cardíaca máxima alcanzada	FrecuenciaMaxima	Numérica
Angina inducida por el ejercicio	AnginaEjercicio	Categórica
Depresión de ST inducida por el ejercicio en relación al reposo	DepresionEjercicioReposo	Numérica
La pendiente del ejercicio pico.	PendienteEjercicioPico	Categórica
Cardiopatía	Cardiopatia	Categórica

El conjunto de datos utilizado en este estudio fue sometido a un preprocesamiento para garantizar la calidad, consistencia y validez de los datos antes de su análisis (Dhaka et al., 2024). Este proceso incluyó la transformación de variables numéricas a formato lógico, la segmentación de datos basada en características específicas y el manejo adecuado de valores faltantes o

inconsistentes (Verma & Khatoon, 2024). La variable cardiopatía, de tipo numérico y con valores de 0 y 1 (donde 1 indica la presencia de insuficiencia cardíaca y 0 su ausencia), fue transformada en una nueva variable denominada Cardioaptia_Log, que representa esta información en formato lógico (verdadero/falso). Esta modificación facilita su integración en análisis categóricos y mejora la claridad en la interpretación de los resultados (Aji et al., 2024). Asimismo, se realizó un procedimiento similar con la variable de azúcar en sangre en ayunas, que clasifica los niveles de glucosa en ayunas según un umbral de 120 mg/dl. Se creó la variable AzucarSangreAyunas_Log, que mantiene el mismo contenido, pero en formato lógico, optimizando su uso en análisis comparativos.

Para asegurar la validez de los datos, se generaron subconjuntos que excluyen registros con valores iguales a 0 en las variables de colesterol, la presión arterial en reposo y la frecuencia cardíaca máxima alcanzada, eliminando posibles inconsistencias o datos incompletos (Sharma & Agarwal, 2024). Esto permitió garantizar que las observaciones utilizadas en el análisis reflejaran información clínica relevante. Además, se implementó una segmentación basada en género, creando subconjuntos específicos para las variables mencionadas, diferenciando entre hombres y mujeres. Estos subconjuntos incluyeron valores de colesterol, presión arterial en reposo y frecuencia cardíaca máxima, lo que permitió analizar diferencias en los patrones clínicos según el género, considerando las variaciones biológicas y fisiológicas asociadas (Vinay et al., 2024).

Finalmente, se abordaron los valores faltantes o inconsistentes mediante la exclusión de celdas vacías en las variables principales, asegurando que el análisis se realizara exclusivamente sobre datos válidos (Vanami et al., 2024). Este enfoque garantizó la integridad de los resultados y minimizó el sesgo potencial derivado de datos incompletos. En conjunto, estas etapas de preprocesamiento optimizaron el conjunto de datos para un análisis robusto y confiable, estableciendo una base sólida para interpretar patrones y tendencias relevantes en el contexto de la insuficiencia cardíaca. La Tabla 2 presenta la descripción de las columnas del conjunto de datos preprocesado.

Tabla 2
Descripción de conjunto de datos preprocesados

Variable	Título	Tipo
Edad	Edad	Numérica
Sexo	Sexo	Categórica
Tipo de dolor en el pecho	TipoDolorPecho	Categórica
Presión arterial en reposo	PresionArterialReposo	Numérica
Colesterol sérico	Colesterol	Numérica
Azúcar en sangre en ayunas	AzucarSangreAyunas_Log	Lógica
Resultados del electrocardiograma en reposo	ECGReposo	Categórica
Frecuencia cardíaca máxima alcanzada	FrecuenciaMaxima	Numérica

Angina inducida por el ejercicio	AnginaEjercicio	Categórica
Depresión de ST inducida por el ejercicio en relación al reposo	DepresionEjercicioReposo	Numérica continua
La pendiente del ejercicio pico.	PendienteEjercicioPico	Categórica
Cardiopatía	Cardiopatía_Log	Lógica

Formulación de Preguntas de Investigación Basadas en las Variables Clínicas

Una vez preprocesado el conjunto de datos, se estableció un conjunto de preguntas clave con el objetivo de investigar cómo las variables clínicas presentes en el conjunto de datos influyen en la probabilidad de insuficiencia cardíaca (Tiwari et al., 2024). Estas preguntas están diseñadas para explorar las relaciones entre las variables, identificar patrones relevantes y profundizar en la comprensión de los factores de riesgo cardiovasculares y representan una guía para realizar análisis estadísticos y visualizaciones dirigidas, con el propósito de responder a las preguntas y respaldar los hallazgos con evidencia cuantitativa. A continuación, se presentan las preguntas de investigación basadas en las variables clínicas (Ayshwarya et al., 2024):

1. ¿Existe una relación entre el sexo y una mayor probabilidad de insuficiencia cardíaca?
2. ¿Existe una relación entre la edad y una mayor probabilidad de insuficiencia cardíaca?
3. ¿Existe relación entre el dolor torácico con una mayor probabilidad de insuficiencia cardíaca?
4. ¿Existe alguna relación entre la presión arterial en reposo con una mayor probabilidad de insuficiencia cardíaca?
5. ¿Existe alguna relación entre el colesterol y una mayor probabilidad de insuficiencia cardíaca?
6. ¿Existe alguna relación entre la glucemia en ayunas y una mayor probabilidad de insuficiencia cardíaca?
7. ¿Existe alguna relación entre el ECG en reposo y una mayor probabilidad de insuficiencia cardíaca?
8. ¿Existe alguna relación entre la frecuencia cardíaca máxima y una mayor probabilidad de insuficiencia cardíaca?
9. ¿Existe alguna relación entre la angina de esfuerzo con una mayor probabilidad de insuficiencia cardíaca?
10. ¿Existe alguna relación entre la depresión de ST inducida por el ejercicio en relación al reposo con una mayor probabilidad de insuficiencia cardíaca?
11. ¿Existe alguna relación entre la pendiente del pico de ejercicio con una mayor probabilidad de insuficiencia cardíaca?
12. ¿Cuál es la combinación de factores de riesgo asociada a una mayor probabilidad de insuficiencia cardíaca?

Visualización de datos

El diseño de la visualización de datos es esencial en el análisis exploratorio de datos, ya que permite identificar patrones, relaciones y tendencias de manera intuitiva y comprensible (Höhn et al., 2024). En esta sección, se describen las estrategias y principios utilizados para desarrollar gráficos claros y precisos basados en las variables del conjunto de datos. Debido a la naturaleza mixta de las variables disponibles (numéricas, categóricas y lógicas), se seleccionaron diferentes tipos de gráficos con la finalidad de representar de una forma óptima cada tipo de dato y las relaciones entre ellos (Nasr et al., 2024). Este diseño sigue estándares y principios de percepción visual asegurando que las gráficas transmitan información relevante sin perder ni distorsionar los datos (Meyer et al., 2024).

El conjunto de datos preprocesado contiene una combinación de variables numéricas, categóricas y lógicas, lo que requiere un enfoque diferenciado para seleccionar los tipos de gráficos más apropiados para su análisis. Se utilizó un diagrama de cajas y bigotes, que se presenta en la figura 1, para visualizar la distribución de la cardiopatía según el sexo y la edad. El histograma que se presenta en la figura 2 fue seleccionado para visualizar la distribución de la variable numérica de la edad, permitiendo identificar patrones en la distribución etaria de los pacientes. Se utilizaron gráficos de barras para comparar las variables categóricas, como el tipo de dolor en el pecho y los resultados del electrocardiograma en reposo, facilitando la comparación de frecuencias entre sus diferentes categorías que se presentan en las figuras 3,4,5,6,7 y 8. Se utilizó el gráfico de crestas de densidad para analizar la relación entre variables numéricas continuas (como el colesterol o la frecuencia cardíaca máxima alcanzada) y variables categóricas (cardiopatía), proporcionando una visualización clara de las distribuciones superpuestas por grupo que se presenta en la figura 9. Finalmente se utilizaron diagramas de dispersión para explorar la relación entre variables numéricas, como la edad y otras características clínicas (por ejemplo, la depresión de ST inducida por el ejercicio con relación al reposo o la presión arterial en reposo), ayudando a identificar tendencias o correlaciones que se presentan en las figuras 10, 11, 12 y 13.

Adicional al tipo de gráfico seleccionado se siguieron los principios técnicos para priorizar el uso de tareas perceptivas fundamentales, como la posición en una escala común (de mayor precisión en la interpretación) y la longitud de elementos gráficos, optimizando la comprensión visual de los datos (Hernly et al., 2024). Se eliminaron elementos visuales innecesarios, como imágenes superfluas o etiquetas redundantes, y se utilizaron colores neutros para evitar distracciones. Las escalas comunes se mantuvieron consistentes en gráficos comparativos, asegurando una representación uniforme. Los gráficos se diseñaron para mantener un factor de mentira cercano a 1, garantizando que las proporciones visuales reflejaran con precisión los valores reales de los datos, evitando exageraciones o minimizaciones. La mayoría de los gráficos incluyeron tres dimensiones para representar simultáneamente la edad, la

cardiopatía y otras variables clínicas, maximizando la cantidad de información transmitida sin comprometer la claridad (Ning et al., 2024). Se optimizó la relación tinta-datos mediante la eliminación de líneas, etiquetas y elementos gráficos innecesarios, enfocando la atención en la información relevante. Esto mejora la legibilidad y minimiza el uso de elementos decorativos. Los gráficos se diseñaron con una proporción visual de 1:1.6, asegurando una presentación balanceada y fácil de interpretar (Tavani et al., 2024). A continuación, se presentan los gráficos realizados en base al diseño planteado arriba.

Figura 1

Diagrama de caja de la distribución del conjunto de datos (verde significa verdadero y amarillo falso)

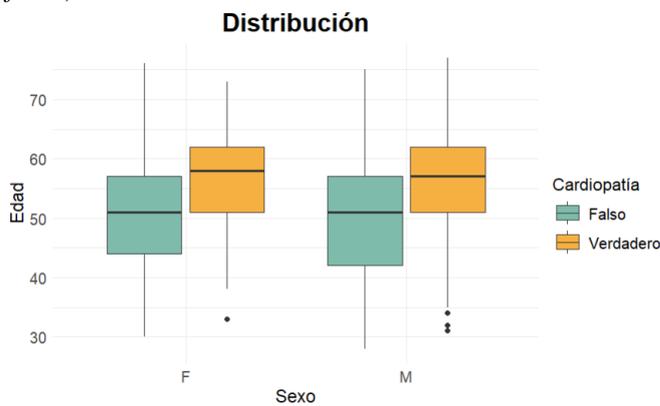


Figura 2

Histograma de la distribución de edad

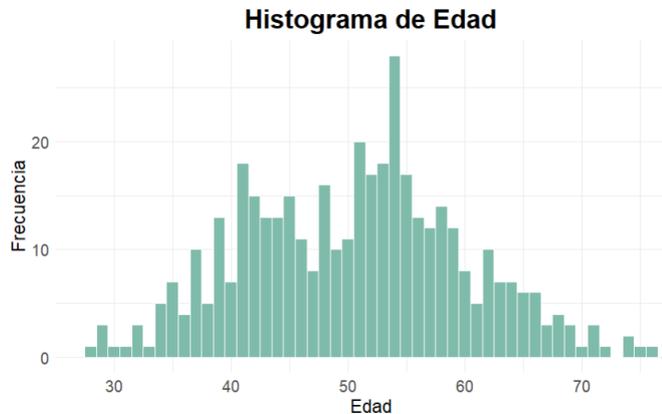


Figura 3

Diagrama de barras de la existencia de cardiopatía según el sexo

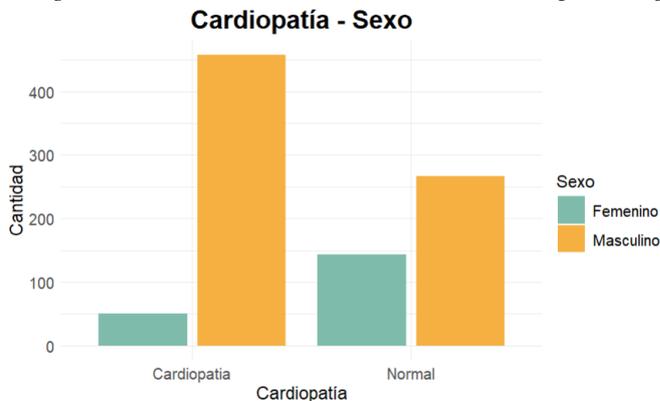


Figura 4

Diagrama de barras de la pendiente del ejercicio pico
Cardiopatía - Pendiente ejercicio pico

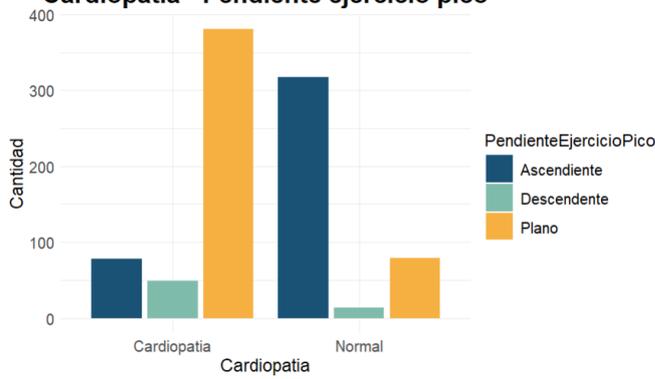


Figura 5

Diagrama de barras de cardiopatía según el ECG en reposo.
Cardiopatía - ECG Reposo

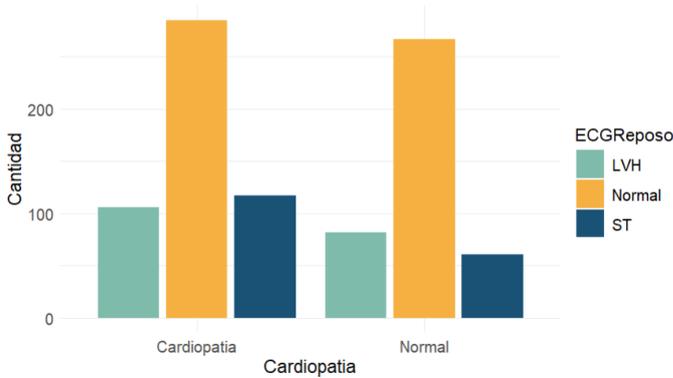


Figura 6

Diagrama de barras de cardiopatía según la angina inducida en el ejercicio.
Cardiopatía - Angina ejercicio

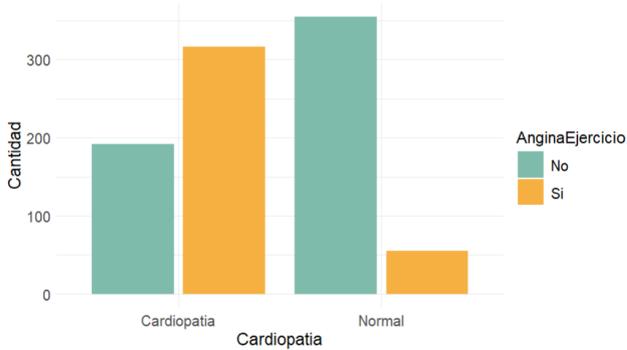


Figura 7

Diagrama de barras de cardiopatía según el tipo de dolor de pecho.

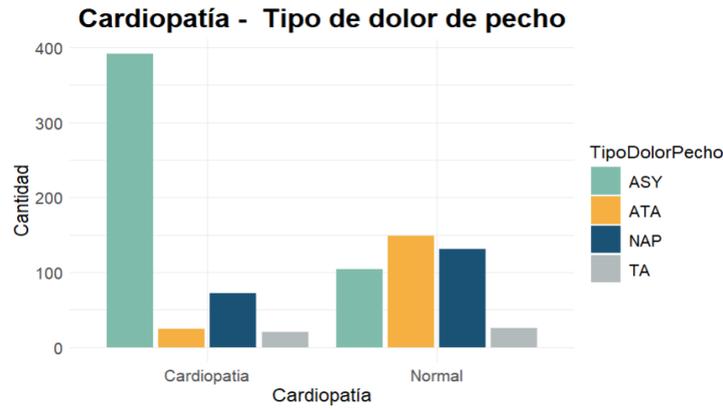


Figura 8

Diagrama de barras de cardiopatía según el azúcar en la sangre en ayunas.

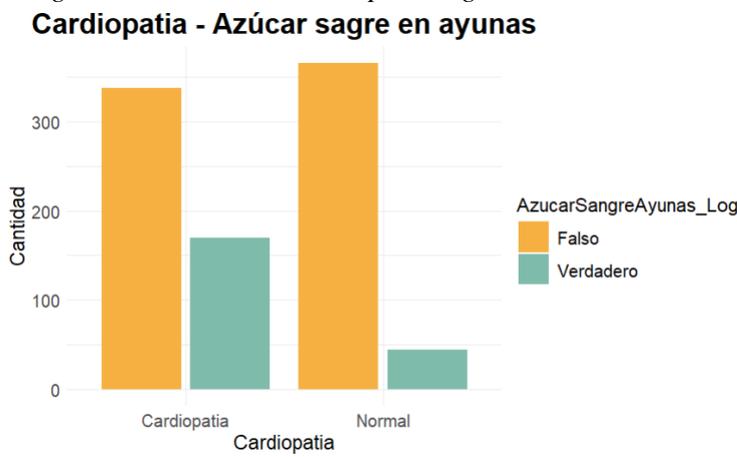


Figura 9

Gráfico de densidad de crestas de cardiopatía según el azúcar en la sangre en ayunas.

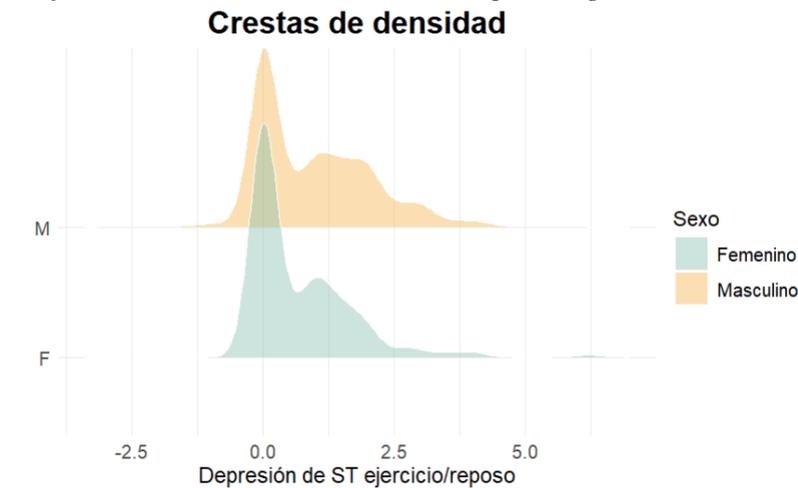


Figura 10

Gráfico de dispersión de cardiopatía según la frecuencia cardiaca máxima en mujeres por edad
Edad mujer - Frecuencia cardiaca máxima

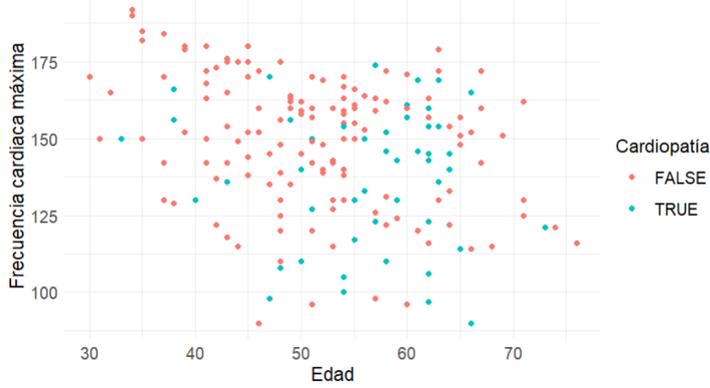


Figura 11

Gráfico de dispersión de cardiopatía según la frecuencia cardiaca máxima en hombres por edad.

Edad hombre - Frecuencia cardiaca máxima

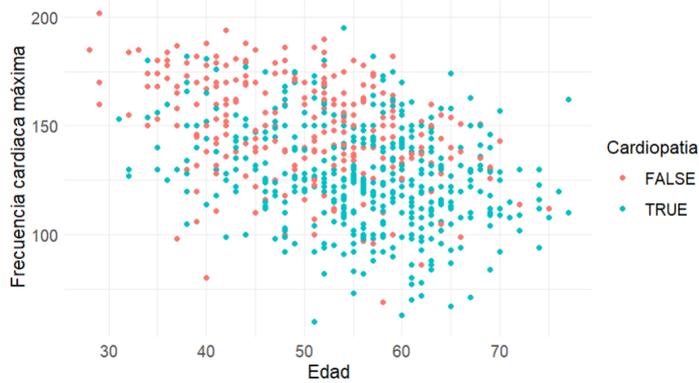


Figura 12

Gráfico de dispersión de cardiopatía según el colesterol en mujeres por edad.
Edad - Colesterol Mujer

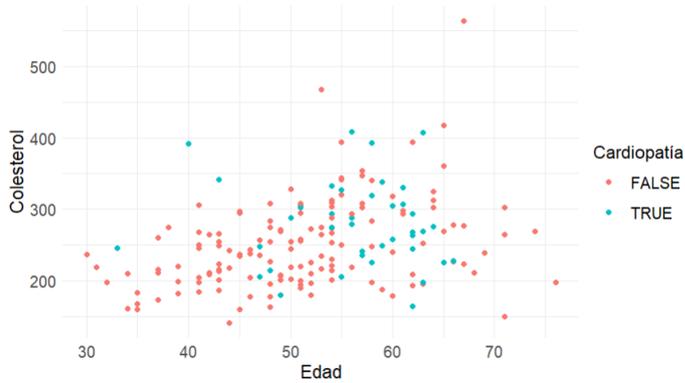
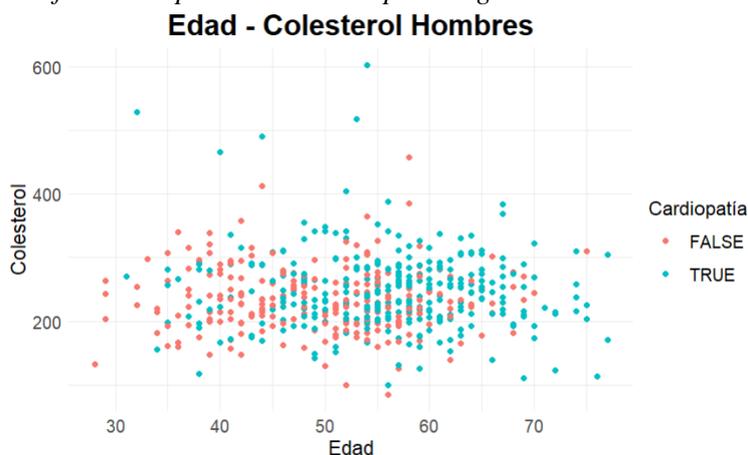


Figura 13

Gráfico de dispersión de cardiopatía según el colesterol en hombres por edad.



RESULTADOS Y DISCUSIÓN

El análisis detallado de las gráficas que representan el conjunto de permitió responder las preguntas planteadas previamente, proporcionando una visión clara de los factores asociados con la probabilidad de desarrollar insuficiencia cardíaca. Los resultados obtenidos se fundamentan en la exploración de datos y la identificación de patrones mediante gráficos y estadísticas descriptivas, y se describen a continuación.

Relación entre el género y la probabilidad de enfermedad cardíaca

Los resultados muestran una clara diferencia en la probabilidad de enfermedad cardíaca entre hombres y mujeres. Los hombres tienen una mayor predisposición a desarrollar insuficiencia cardíaca en comparación con las mujeres. Este hallazgo resalta el género como un factor de riesgo importante.

Relación entre la edad y la probabilidad de enfermedad cardíaca

La edad se confirma como un factor relevante en la aparición de enfermedades cardíacas. Los pacientes mayores de 55 años presentan un riesgo significativamente mayor de desarrollar insuficiencia cardíaca.

Relación entre el tipo de dolor torácico y la probabilidad de enfermedad cardíaca

El tipo de dolor de pecho también tiene un impacto significativo. Los pacientes con angina atípica tienen una mayor probabilidad de presentar insuficiencia cardíaca.

Relación entre la presión arterial en reposo y la probabilidad de enfermedad cardíaca

La presión arterial en reposo mostró una relación interesante, especialmente en mujeres. Aquellas con valores elevados de presión arterial en reposo tienen una mayor probabilidad de desarrollar insuficiencia cardíaca.

Relación entre el colesterol y la probabilidad de enfermedad cardíaca

Contrario a las expectativas, los niveles de colesterol no se asociaron significativamente con un mayor riesgo de insuficiencia cardíaca en este análisis.

Relación entre la glucosa en ayunas y la probabilidad de enfermedad cardíaca.

Los resultados indican que los niveles de glucosa en ayunas tampoco influyen significativamente en la probabilidad de insuficiencia cardíaca.

Relación entre el electrocardiograma en reposo y la probabilidad de enfermedad cardíaca

El análisis del electrocardiograma en reposo no reveló una relación significativa con la insuficiencia cardíaca.

Relación entre la frecuencia cardíaca máxima y la probabilidad de enfermedad cardíaca

La frecuencia cardíaca máxima alcanzada durante el esfuerzo no mostró ser un factor significativo en el desarrollo de insuficiencia cardíaca.

Relación entre la angina inducida por ejercicio y la probabilidad de enfermedad cardíaca

La angina inducida por ejercicio sí presentó una asociación significativa. Los pacientes que experimentaron dolor torácico durante el ejercicio tienen una mayor probabilidad de presentar insuficiencia cardíaca, lo que refuerza la importancia de incluir esta variable en evaluaciones clínicas de riesgo.

Relación entre el indicador la depresión del ST inducida durante el ejercicio en relación al reposo y la probabilidad de enfermedad cardíaca

El indicador de la depresión del ST inducida durante el ejercicio en relación al reposo no mostró una influencia relevante en la probabilidad de insuficiencia cardíaca, sugiriendo que este factor puede no ser un marcador decisivo en este contexto.

Relación entre la pendiente del ejercicio pico y la probabilidad de enfermedad cardíaca

Los pacientes con una pendiente del ejercicio pico tienen un riesgo significativamente mayor de insuficiencia cardíaca.

Combinación de factores de riesgo asociados a una mayor probabilidad de insuficiencia cardíaca

Se identificó que los hombres mayores de 55 años con dolor torácico atípico (ASY) y angina inducida por ejercicio son los más propensos a desarrollar insuficiencia cardíaca. Este conjunto de características representa un perfil de alto riesgo que puede ser utilizado para priorizar intervenciones clínicas y estrategias de prevención.

Se identificó que ciertas variables demográficas y clínicas tienen un impacto significativo en la probabilidad de insuficiencia cardíaca. Primero, se identificó que el género y la edad se identificaron como factores relevantes, con los hombres y los pacientes mayores de 55 años mostrando un mayor riesgo de desarrollar la enfermedad.

El análisis del tipo de dolor torácico evidenció que los pacientes con angina atípica (ASY) tienen una probabilidad significativamente mayor de padecer insuficiencia cardíaca, lo que posiciona esta variable como un indicador clínico crítico en la evaluación inicial de riesgos. De manera similar, la angina inducida por ejercicio mostró una asociación fuerte con la insuficiencia

cardíaca, reforzando su papel como un marcador clave en la detección temprana de problemas cardiovasculares.

Por otro lado, variables como el colesterol, el nivel de azúcar en ayunas, la frecuencia cardíaca máxima y el electrocardiograma en reposo no mostraron una influencia estadísticamente relevante en el desarrollo de la enfermedad dentro del contexto de este conjunto de datos. Esto sugiere que estos factores, aunque importantes en otros escenarios, podrían tener una menor sensibilidad o especificidad en la predicción de insuficiencia cardíaca en esta muestra específica. La pendiente del ejercicio pico se destacó como un factor crítico, con pacientes que presentan una pendiente plana mostrando un mayor riesgo de insuficiencia cardíaca. Este hallazgo resalta la importancia de incluir medidas relacionadas con la pendiente del ejercicio pico en la evaluación clínica, especialmente en pruebas de esfuerzo.

En base a este análisis se puede determinar que existe una combinación de factores de riesgo asociados a una mayor probabilidad de insuficiencia cardíaca.

CONCLUSIONES

El presente estudio analizó el conjunto de datos de predicción de insuficiencia cardíaca mediante técnicas de análisis y visualización de datos para identificar los factores clínicos más relevantes asociados a la probabilidad de insuficiencia cardíaca. Los resultados obtenidos permiten extraer conclusiones importantes tanto en términos clínicos como metodológicos, destacando la utilidad del enfoque analítico aplicado.

Se identificó un perfil de alto riesgo compuesto por hombres mayores de 55 años con angina atípica y dolor torácico inducido por ejercicio. Este perfil clínico puede ser utilizado para priorizar la atención médica y diseñar estrategias de intervención dirigidas a mitigar el riesgo en poblaciones específicas.

Desde un enfoque metodológico, este estudio demuestra la importancia de seleccionar visualizaciones apropiadas y aplicar principios teóricos de representación gráfica para maximizar la comprensión de los datos y garantizar la precisión de las interpretaciones. El uso de gráficos específicos para representar variables categóricas, numéricas y lógicas permitió abordar preguntas clínicas clave y extraer conclusiones significativas.

Este análisis reafirma la relevancia de un enfoque basado en datos para identificar factores de riesgo y perfilar grupos poblacionales en la prevención y manejo de enfermedades cardiovasculares. Además, establece una base sólida para el desarrollo de modelos predictivos más robustos que integren variables críticas y mejoren la detección temprana de insuficiencia cardíaca. Los hallazgos obtenidos pueden ser útiles tanto en la práctica clínica como en la investigación, contribuyendo al avance en el entendimiento de esta patología y sus factores asociados.

Agradecimientos

Los autores desean expresar su más profundo agradecimiento a la red de investigación INTELIA, respaldada por REDU, por su invaluable apoyo y colaboración a lo largo del desarrollo de este trabajo.

REFERENCIAS

- Aghapanah, H., Rasti, R., Tabesh, F., Pouraliakbar, H., Sanei, H., & Kermani, S. (2025). MECardNet: A novel multi-scale convolutional ensemble model with adaptive deep supervision for precise cardiac MRI segmentation. *Biomedical Signal Processing and Control*, *100*, 106919. <https://doi.org/10.1016/j.bspc.2024.106919>
- Aji, N. B., Kurnianingsih, K., Masuyama, N., & Nojima, Y. (2024). CNN-LSTM for Heartbeat Sound Classification. *JOIV : International Journal on Informatics Visualization*, *8*(2), 735. <https://doi.org/10.62527/joiv.8.2.2115>
- Ayshwarya, B., George, A., & M, D. (2024). *Heart Failure Prediction for a Patient Using Hybrid African Buffalo Optimization with Naive Bayes Machine Learning Techniques* (pp. 141–154). https://doi.org/10.1007/978-981-99-8612-5_12
- Dhaka, P., Sehrawat, R., & Bhutani, P. (2024). An efficient heart disease prediction model using particle swarm–optimized ensemble classifier model. In *Securing Next-Generation Connected Healthcare Systems* (pp. 123–135). Elsevier. <https://doi.org/10.1016/B978-0-443-13951-2.00005-2>
- Garai, S., Kashyap, P., Irfan, S., & Pareek, A. (2024). Cardiac Disease Risks Pattern Recognition Using Advanced Predictive Analytics. In *Healthcare Analytics and Advanced Computational Intelligence* (pp. 164–179). CRC Press. <https://doi.org/10.1201/9781032624891-12>
- Hernly, E., Hu, H., & Laskin, J. (2024). MSIGen: An Open-Source Python Package for Processing and Visualizing Mass Spectrometry Imaging Data. *Journal of the American Society for Mass Spectrometry*, *35*(10), 2315–2323. <https://doi.org/10.1021/jasms.4c00178>
- Höhn, M., Schwindt-Drews, S., Hahn, S., Patyna, S., Büttner, S., & Kohlhammer, J. (2024). RenalViz: Visual analysis of cohorts with chronic kidney disease. *Computers & Graphics*, *125*, 104120. <https://doi.org/10.1016/j.cag.2024.104120>
- Ijuin, S., Inoue, A., Hifumi, T., Taira, T., Suga, M., Nishimura, T., Sakamoto, T., Kuroda, Y., & Ishihara, S. (2025). Analysis of factors associated with favorable neurological outcomes in patients with initial PEA who underwent ECPR - A secondary analysis of the SAVE-J II study. *Journal of Critical Care*, *85*, 154917. <https://doi.org/10.1016/j.jcrc.2024.154917>
- Li, H., & Xia Dou, Y. (2025). Resource optimization in smart electronic health systems using IoT for heart disease prediction via feedforward neural networks. *Cluster Computing*, *28*(1), 21. <https://doi.org/10.1007/s10586-024-04726-7>
- Meyer, L., Eling, N., & Bodenmiller, B. (2024). cytoviewer: an R/Bioconductor package for interactive visualization and exploration of highly multiplexed imaging data. *BMC Bioinformatics*, *25*(1), 9. <https://doi.org/10.1186/s12859-023-05546-z>

- Nasr, H. A., Aref, N.-E. M., Ellah, M. R. A., & Abdelhakiem, M. A. H. (2024). Cardiac biomarkers as tools in the prediction and diagnosis of traumatic pericarditis and traumatic reticuloperitonitis in cattle and buffaloes. *BMC Veterinary Research*, 20(1), 329. <https://doi.org/10.1186/s12917-024-04174-w>
- Ning, J., Zhao, H., & Xu, X. (2024). A volume rendering visualization method oriented to explosion field features. *AIP Advances*, 14(9). <https://doi.org/10.1063/5.0226283>
- Sharma, P., & Agarwal, S. (2024). Cardiovascular Disease Analysis Using Different Machine Learning Techniques. *2024 2nd International Conference on Artificial Intelligence and Machine Learning Applications Theme: Healthcare and Internet of Things (AIMLA)*, 1–6. <https://doi.org/10.1109/AIMLA59606.2024.10531431>
- Tavani, S., Corradetti, A., Rizzo, R. E., & Seers, T. D. (2024). Best practices towards the digitization of 3D traces from virtual outcrop models. *Journal of Structural Geology*, 186, 105222. <https://doi.org/10.1016/j.jsg.2024.105222>
- Tiwari, S., Singh, S., & Tiwari, D. (2024). Comparative Strategies for Anticipating Cardiovascular Maladies: An In-Depth Analytical Interpretation. *2024 2nd International Conference on Disruptive Technologies (ICDT)*, 981–985. <https://doi.org/10.1109/ICDT61202.2024.10489512>
- Vanami, N. V. J., Chintalapati, L. R., Challagundla, Y., & Mohanty, S. N. (2024). *Feature Selection Using Data Mining Techniques for Prognostication of Cardiovascular Diseases* (pp. 344–353). https://doi.org/10.1007/978-3-031-50571-3_24
- Verma, P., & Khatoun, A. (2024). Data Mining Applications in Healthcare: A Comparative Analysis of Classification Techniques for Diabetes Diagnosis Using the PIMA Indian Diabetes Dataset. *2024 4th International Conference on Innovative Practices in Technology and Management (ICIPTM)*, 1–5. <https://doi.org/10.1109/ICIPTM59628.2024.10563296>
- Vinay, N. A., Vidyasagar, K. N., Rohith, S., Pruthviraja, D., Supreeth, S., & Bharathi, S. H. (2024). An RNN-Bi LSTM Based Multi Decision GAN Approach for the Recognition of Cardiovascular Disease (CVD) From Heart Beat Sound: A Feature Optimization Process. *IEEE Access*, 12, 65482–65502. <https://doi.org/10.1109/ACCESS.2024.3397574>