

UAX

TRABAJO FIN DE MÁSTER

**MUIA
24-25**

UNIVERSIDAD ALFONSO X EL SABIO

Business Tech

Máster Universitario en Inteligencia Artificial



TRABAJO DE FIN DE MÁSTER

**Diseño e implementación de modelos de
Inteligencia Artificial para la identificación
temprana de pacientes con riesgo de sufrir
ataques cardíacos**

JON MAESTRE ESCOBAR
Beatriz Magán Pinto

junio 2025

RESUMEN

Este proyecto tiene como objetivo diseñar e implementar modelos avanzados de Inteligencia Artificial (IA) para la identificación temprana de pacientes con riesgo de sufrir ataques cardíacos, utilizando datos clínicos y biométricos para generar predicciones precisas. A través del análisis exhaustivo de un conjunto de datos recopilado en la India, que incluye información detallada sobre signos vitales, hábitos y antecedentes médicos, el proyecto construye una base sólida para el análisis exploratorio y el modelado predictivo. El estudio comienza con una fase integral de ciencia de datos, que incluye análisis exploratorio y visualización de datos para comprender la distribución y el comportamiento de las variables implicadas, así como su relevancia en la predicción de eventos cardiovasculares.

El proyecto emplea una amplia variedad de algoritmos de machine learning y deep learning, entre los que se incluyen regresión logística, random forest, gradient boosting, redes neuronales artificiales y redes neuronales convolucionales. Además, se contempla la posibilidad de aplicar modelos basados en transformers si la estructura y complejidad de los datos lo permiten. Estos métodos han sido seleccionados por su eficacia en tareas de clasificación y por su capacidad para modelar interacciones no lineales entre factores de riesgo.

El conjunto de datos utilizado en este estudio proviene de datos de salud disponibles públicamente recopilados en la India. No obstante, se ha requerido un importante proceso de limpieza y preparación, que incluyó el tratamiento de valores ausentes, la normalización de variables numéricas, la codificación de variables categóricas y el balanceo de clases en la variable objetivo. La versión final del conjunto de datos fue transformada en un formato limpio y estructurado, apto para el entrenamiento y evaluación de modelos.

El objetivo principal del proyecto es identificar las variables que más influyen en el riesgo de infarto y construir modelos capaces de clasificar con precisión a los pacientes según su nivel de riesgo. Estas predicciones podrían apoyar la toma de decisiones clínicas mediante intervenciones tempranas en pacientes de alto riesgo. Se espera que los modelos basados en deep learning presenten el mejor rendimiento, especialmente por su capacidad de capturar relaciones complejas dentro del conjunto de datos.

En resumen, este proyecto representa un avance significativo en la aplicación de la IA al ámbito de la medicina preventiva, ofreciendo conocimientos valiosos que podrían contribuir a reducir la mortalidad por enfermedades cardiovasculares mediante una detección del riesgo más temprana y precisa.

Resumen previo. Al finalizar, lo reescribiré adaptando los cambios.

Palabras clave

Inteligencia Artificial, Salud Cardiovascular, Machine Learning, Deep Learning, Modelos Predictivos.

ABSTRACT

This project aims to design and implement advanced Artificial Intelligence (AI) models for the early identification of patients at risk of suffering heart attacks, leveraging clinical and biometric data to generate accurate predictions. By thoroughly analyzing a dataset collected from patients in India, which includes detailed information on vital signs, habits, and medical history, the project builds a robust foundation for both data exploration and predictive modeling. The study begins with a comprehensive data science phase, which includes extensive exploratory data analysis and visualization to understand the distribution and behavior of the variables involved, as well as their relevance in the prediction of cardiovascular events.

The project employs a wide range of machine learning and deep learning algorithms, including logistic regression, random forests, gradient boosting machines, artificial neural networks, and convolutional neural networks. Additionally, transformer-based models may be explored if data complexity and structure allow for their effective application. These methods are selected for their proven capabilities in classification tasks and their ability to model nonlinear interactions between risk factors.

The dataset used in this study originates from publicly available health data collected in India. However, it required significant preprocessing to handle missing values, normalize numerical features, encode categorical variables, and balance the target classes. The final version of the dataset was transformed into a clean and structured format suitable for model training and evaluation.

The primary objective of the project is to identify which variables most strongly influence the risk of a heart attack and to build models capable of accurately classifying patients according to their risk level. These predictions could support clinical decision-making by enabling early interventions for high-risk individuals. Among the models tested, those based on deep learning are expected to show superior performance, particularly in capturing complex relationships within the data.

In summary, this project represents a meaningful step forward in applying AI to the field of preventive medicine. It offers valuable insights that may help reduce cardiovascular disease mortality through earlier and more accurate risk detection.

Preliminary summary. When finished, I will rewrite it to adapt the changes

Keywords

Artificial Intelligence, Cardiovascular Health, Machine Learning, Deep Learning, Predictive Modeling.

ÍNDICE

1. INTRODUCCIÓN	9
2. ESTADO DEL ARTE	11
2.1. Contexto	12
2.2. Justificación	13
2.3. Análisis Comparativo de Estudios Existentes	14
2.3.1 Revisión General de Estudios Existentes	15
2.3.2 Metodologías Empleadas en Investigaciones Previas.....	25
2.3.3 Conjuntos de Datos Utilizados en Estudios Relacionados.....	30
2.3.4 Resultados y Hallazgos	33
2.3.5 Casos de Estudio en Proyectos Reales	38
2.3.6 Gráficos y Tabla Comparativa.....	41
2.3.7 Comparaciones con mi Trabajo Fin de Máster	44
3. OBJETIVOS	49
3.1. Objetivos Generales	50
3.2. Objetivos Específicos	50
3.3. Alcance	52
3.4. Limitaciones.....	53
4. METODOLOGÍA	56

4.1. Recopilación y Preprocesamiento de Datos	56
4.2. Desarrollo y Entrenamiento de los Modelos (es hipotético)	57
4.3. Evaluación de los Modelos (es hipotético)	57
4.4. Gestión de Archivos con Google Drive y GitHub (es hipotético aun no los tengo creados pero lo estarán)	58
4.5. Planificación y Organización de Tareas con ClickUp (es hipotético aun no está creado pero lo estará)	59
5. PLANIFICACIÓN Y PRESUPUESTO (es hipotético)	60
5.1. Planificación	60
5.2. Presupuesto	62
6. DESARROLLO DEL PROYECTO	64
7. BIBLIOGRAFÍA	65

ÍNDICE DE ILUSTRACIONES

ILUSTRACIÓN 1: PRECISIÓN DE LOS MODELOS DE IA EN LOS ESTUDIOS ANALIZADOS	43
ILUSTRACIÓN 2: FRECUENCIA DE PREDICTORES CLAVE EN LOS ESTUDIOS ANALIZADOS	43

ÍNDICE DE TABLAS

TABLA 1: COMPARATIVA DE ASPECTOS CLAVE DE CADA ESTUDIO	42
--	----

1. INTRODUCCIÓN

En los últimos años, la integración de la Inteligencia Artificial (IA) en diversos sectores ha generado transformaciones significativas, especialmente en el ámbito de la salud. La capacidad de los modelos de IA para procesar grandes volúmenes de datos y extraer patrones relevantes ha abierto nuevas posibilidades para el diagnóstico, la prevención y la personalización de tratamientos médicos. Dentro de este contexto, este proyecto se centra en el diseño e implementación de modelos avanzados de IA para la identificación temprana de pacientes con riesgo de sufrir ataques cardíacos, haciendo uso de datos clínicos y biométricos recogidos en la India.

El conjunto de datos utilizado ha requerido un proceso exhaustivo de preprocesamiento con el fin de consolidar una base de datos limpia, coherente y estructurada, adecuada para la aplicación de técnicas de ciencia de datos y modelado predictivo. Inicialmente, los datos presentaban inconsistencias, valores ausentes y una estructura poco eficiente para el análisis. Una vez finalizado el preprocesamiento, se ha obtenido un dataset óptimo para el desarrollo de modelos de machine learning y deep learning.

Comprender los factores que influyen en la aparición de eventos cardiovasculares es esencial para desarrollar herramientas de intervención temprana que puedan salvar vidas. Las técnicas tradicionales han mostrado limitaciones a la hora de identificar relaciones no lineales o complejas entre múltiples variables clínicas. Por ello, este proyecto emplea algoritmos como regresión logística, random forest y redes neuronales, además de explorar modelos avanzados como redes convolucionales y, si la naturaleza de los datos lo permite, transformers. Estas metodologías permiten capturar patrones complejos y realizar predicciones precisas sobre el riesgo cardíaco de los pacientes.

El objetivo principal de esta investigación es identificar los factores clave que determinan el riesgo de infarto y desarrollar modelos capaces de clasificarlos

con alta precisión. De este modo, se busca facilitar a los profesionales sanitarios la toma de decisiones proactivas y personalizadas, mejorando así la eficacia de las medidas preventivas. Este trabajo contribuye además al campo de la salud digital, proporcionando un enfoque basado en datos que puede ser escalado e implementado en sistemas de atención médica.

Los datos utilizados en este estudio corresponden a registros recopilados tras el inicio de la pandemia por COVID-19, un acontecimiento que ha puesto de manifiesto la necesidad de reforzar la prevención y el monitoreo de enfermedades crónicas, como las cardiovasculares. La pandemia ha evidenciado la importancia de disponer de sistemas inteligentes capaces de anticiparse a riesgos de salud mediante herramientas tecnológicas eficaces. En este sentido, el presente proyecto pretende cubrir esta necesidad, desarrollando un modelo predictivo que pueda ser utilizado en distintos contextos médicos para mejorar la detección temprana de riesgos cardíacos.

En cuanto al resultado esperado, se busca construir un modelo robusto capaz de ofrecer predicciones fiables y fácilmente interpretables para su uso en entornos clínicos. La implementación de este tipo de soluciones podría revolucionar los sistemas de diagnóstico preventivo, estableciendo un nuevo estándar en el uso de la IA para mejorar la salud cardiovascular y, en general, la calidad de vida de los pacientes.

En resumen, este proyecto representa un avance significativo en la aplicación de la inteligencia artificial en el ámbito de la medicina preventiva. Al establecer nuevos enfoques en la personalización del diagnóstico, se aspira a contribuir a la mejora de los sistemas de salud y a la reducción de la mortalidad asociada a enfermedades cardíacas.

2. ESTADO DEL ARTE

El avance de la Inteligencia Artificial ha permitido el desarrollo de herramientas cada vez más precisas para el análisis de datos médicos y la predicción de enfermedades. En particular, el uso de técnicas de machine learning y deep learning en el ámbito de la salud ha cobrado una gran relevancia en los últimos años, especialmente en la prevención de enfermedades cardiovasculares, que continúan siendo una de las principales causas de mortalidad a nivel mundial. La posibilidad de detectar de forma temprana a pacientes con riesgo de sufrir un ataque cardíaco resulta fundamental para aplicar medidas preventivas que reduzcan tanto la carga asistencial como el impacto sobre la calidad de vida de los pacientes.

Este apartado tiene como finalidad presentar una revisión de los trabajos más relevantes relacionados con la aplicación de modelos de IA en la predicción del riesgo cardiovascular. Se analizarán investigaciones recientes que han empleado diferentes técnicas de aprendizaje supervisado, redes neuronales profundas y arquitecturas especializadas como los transformers, así como estudios centrados en el análisis de variables clínicas y biométricas mediante enfoques de ciencia de datos. Asimismo, se revisarán las principales fuentes de datos utilizadas en este tipo de investigaciones, los métodos de evaluación aplicados y los desafíos comunes en cuanto a la calidad, disponibilidad y representatividad de los datos.

A través de esta revisión, se busca contextualizar el presente proyecto dentro del panorama actual de la investigación en salud digital, identificar las principales líneas de trabajo existentes y justificar la elección metodológica adoptada. Esta base teórica servirá como punto de partida para el diseño y la implementación de los modelos desarrollados en este trabajo.

2.1. Contexto

A pesar de los avances en medicina preventiva y del creciente uso de tecnologías orientadas al diagnóstico temprano, las enfermedades cardiovasculares continúan siendo una de las principales causas de mortalidad en todo el mundo. En particular, los ataques cardíacos representan un desafío urgente para los sistemas de salud debido a su alta prevalencia y al hecho de que muchos casos podrían haberse evitado con una detección precoz y un seguimiento adecuado de los factores de riesgo. Uno de los principales problemas radica en la dificultad para identificar a tiempo a aquellos pacientes con mayor probabilidad de sufrir un evento cardiovascular, especialmente cuando los síntomas no se manifiestan de forma clara o directa.

En este sentido, la Inteligencia Artificial se presenta como una solución prometedora. Los modelos avanzados de IA, especialmente aquellos basados en técnicas de aprendizaje automático y aprendizaje profundo, ofrecen la posibilidad de analizar grandes volúmenes de datos clínicos, detectar patrones complejos y realizar predicciones precisas sobre el estado de salud de los pacientes. Sin embargo, para implementar eficazmente estos modelos, es fundamental comprender qué variables influyen en el desarrollo de enfermedades cardíacas y desarrollar algoritmos que se adapten a la naturaleza y la variabilidad de los datos médicos.

Además, tras la pandemia de COVID-19, la atención hacia la salud cardiovascular ha cobrado aún mayor relevancia. El impacto del virus en personas con afecciones cardíacas preexistentes ha evidenciado la necesidad de contar con herramientas tecnológicas que permitan anticiparse a este tipo de situaciones y fortalecer las estrategias de prevención. Por tanto, surge una oportunidad para aprovechar el potencial de la IA en el análisis de datos clínicos con el fin de mejorar la predicción y gestión del riesgo cardíaco.

Por otra parte, diversos estudios recientes han demostrado que el uso de modelos predictivos basados en IA mejora la capacidad de detección temprana

y permite una intervención más personalizada y oportuna. Al aplicar algoritmos a conjuntos de datos de salud, se pueden identificar con mayor precisión aquellos factores que incrementan el riesgo de infarto, lo cual resulta fundamental para diseñar planes de actuación adaptados a cada paciente. Este nivel de personalización es clave para optimizar los recursos sanitarios y reducir la incidencia de eventos cardiovasculares graves.

Este proyecto se enfoca en integrar modelos avanzados de Inteligencia Artificial al análisis de datos clínicos con el propósito de identificar patrones de riesgo en pacientes y facilitar la toma de decisiones médicas. A través del estudio de un conjunto de datos procedente de la India, se pretende analizar qué variables resultan más relevantes en la predicción del riesgo cardíaco, con el objetivo de ofrecer un modelo fiable, escalable y útil en contextos clínicos reales.

2.2. Justificación

La implementación de modelos de Inteligencia Artificial en el ámbito sanitario no solo tiene el potencial de mejorar la detección temprana de enfermedades cardiovasculares, sino que también puede suponer una contribución significativa al campo de la salud digital. Al proporcionar recomendaciones basadas en datos y facilitar la identificación de pacientes en situación de riesgo, se promueve una medicina más proactiva, personalizada y eficiente. Esto implica un beneficio directo para cada individuo, pero también un impacto positivo a nivel de salud pública, al reducir la incidencia de eventos cardiovasculares graves y optimizar los recursos del sistema sanitario.

Los algoritmos de IA permiten realizar un análisis detallado y continuo de los datos clínicos, lo que facilita la identificación de patrones y correlaciones que difícilmente podrían ser detectados mediante métodos tradicionales. Además, aunque este proyecto no contempla el análisis en tiempo real, la posibilidad de aplicar estos modelos en entornos dinámicos permitiría adaptar los planes

de seguimiento clínico a las necesidades cambiantes de cada paciente, incrementando así su eficacia.

Asimismo, la integración de IA en procesos de diagnóstico y prevención puede dar lugar al desarrollo de estrategias más avanzadas para la gestión del riesgo cardiovascular. Por ejemplo, mediante el análisis predictivo, es posible anticipar qué pacientes presentan una mayor probabilidad de sufrir un infarto, permitiendo una intervención temprana y dirigida. Este enfoque proactivo no solo mejora los resultados clínicos, sino que también contribuye a reducir la carga sobre los profesionales sanitarios y los sistemas de atención médica.

Resumiendo, este proyecto tiene el potencial de establecer un nuevo estándar en la aplicación de la inteligencia artificial en medicina preventiva, al utilizar modelos avanzados para detectar de forma precisa y anticipada el riesgo de sufrir ataques cardíacos. De este modo, no solo se mejora la experiencia del paciente y la toma de decisiones médicas, sino que también se promueve una mayor adopción de soluciones tecnológicas en el ámbito de la salud, contribuyendo al bienestar general de la población.

2.3. Análisis Comparativo de Estudios Existentes

Este apartado presenta una comparativa entre diversos estudios científicos y proyectos reales centrados en la aplicación de la Inteligencia Artificial para la predicción del riesgo de enfermedades cardiovasculares, en particular, ataques cardíacos. El objetivo es mostrar de forma clara el estado actual de la investigación en este ámbito, identificar los avances alcanzados por otros autores y destacar las aportaciones diferenciales que este proyecto propone. Para ello, el análisis se dividirá en varias subcategorías que permitan una revisión exhaustiva y organizada de los trabajos relacionados.

2.3.1 Revisión General de Estudios Existentes

En los últimos años, la aplicación de modelos de Inteligencia Artificial en el campo de la salud ha despertado un creciente interés, especialmente en la predicción de eventos cardiovasculares mediante el análisis de datos clínicos y biométricos. Numerosos estudios han abordado esta problemática desde distintas perspectivas, empleando algoritmos de machine learning y deep learning para anticipar el riesgo cardíaco en pacientes. Esta sección ofrece una visión general de los trabajos más relevantes en este campo, analizando sus metodologías, conjuntos de datos utilizados, resultados obtenidos y su aplicabilidad en entornos reales.

El Rol de la IA en la Predicción de Ataques Cardíacos

La aplicación de la Inteligencia Artificial en la predicción de ataques cardíacos ha evolucionado significativamente en los últimos años, convirtiéndose en una herramienta clave para el diagnóstico temprano y la toma de decisiones clínicas. La disponibilidad de datos médicos provenientes de bases públicas, hospitales e instituciones académicas ha permitido el desarrollo de modelos capaces de identificar patrones complejos que pueden pasar desapercibidos mediante métodos clínicos tradicionales.

Uno de los estudios más relevantes en este ámbito es el de Alexander y Wang (2017), quienes realizaron una revisión sobre el uso de tecnologías de Big Data para la predicción y gestión de enfermedades cardiovasculares. Este trabajo subraya la importancia de integrar técnicas como data mining, Hadoop y visualización de datos a gran escala, destacando cómo estas herramientas pueden personalizar tratamientos médicos y anticipar eventos cardíacos críticos a partir del análisis de datos masivos.

Por otro lado, Takci (2018) abordó la mejora en la predicción de ataques cardíacos mediante la combinación de algoritmos de selección de características y métodos de aprendizaje automático. Usando el dataset

Statlog (Heart), demostró que la integración del algoritmo de soporte vectorial (SVM) con kernel lineal y el método de selección ReliefF alcanzó una precisión del 84,81 %, evidenciando la influencia de una adecuada ingeniería de características en el rendimiento del modelo.

El estudio de Patil y Kumaraswamy (2009) propuso una metodología basada en minería de datos para extraer patrones significativos desde almacenes clínicos. Mediante la combinación de clustering con K-means y el algoritmo MAFLA para la identificación de itemsets frecuentes, se desarrolló un sistema de predicción que resalta el valor del conocimiento oculto en grandes volúmenes de datos médicos.

Alshraideh et al. (2024) exploraron distintos algoritmos de IA aplicados a datos del Hospital Universitario de Jordania. Utilizando clasificadores como Random Forest, SVM y KNN junto con Particle Swarm Optimization (PSO) para seleccionar las variables más relevantes, obtuvieron una precisión destacada del 94,3 % con SVM+PSO. Este resultado refleja el potencial de la IA para alcanzar niveles de precisión competitivos en contextos clínicos reales.

Desde una perspectiva más técnica, Gupta et al. (2021) aplicaron un enfoque de pipeline de aprendizaje automático sobre conjuntos de datos del Framingham Heart Study y del repositorio UCI. Comparando modelos como Gradient Boosting, Árboles de Decisión y Regresión Logística, identificaron al Gradient Boosting como el más eficaz, y destacaron variables como tipo de dolor torácico, colesterol y frecuencia cardíaca como las de mayor influencia en la predicción.

Nandal et al. (2022) desarrollaron el método ML-HAP, centrado en la predicción sintomática de ataques cardíacos mediante modelos como SVM, XGBoost y Regresión Logística. Su estudio mostró que XGBoost ofrecía el mejor rendimiento, con un AUC de 0,94, reafirmando la utilidad de los

algoritmos de boosting en tareas de clasificación médica altamente específicas.

En una línea similar, Feng et al. (2024) propusieron un enfoque centrado en el uso del algoritmo XGBoost, acompañado de una rigurosa selección de características y optimización de hiperparámetros. Su modelo demostró ser altamente interpretable y eficaz, superando a otros métodos tanto en precisión como en escalabilidad, lo cual es esencial para su implementación práctica en sistemas de salud.

Finalmente, Waqar et al. (2021) introdujeron un modelo basado en redes neuronales profundas integradas con la técnica SMOTE para abordar el desbalance de clases en los datos. Su enfoque evitó la necesidad de ingeniería de características manual, lo que lo convierte en una solución eficiente y rentable. Este modelo destacó por su alta fiabilidad en la clasificación, lo que refuerza su aplicabilidad clínica incluso en contextos con recursos limitados.

En conjunto, estos estudios demuestran que la IA no solo puede mejorar significativamente la precisión de las predicciones sobre ataques cardíacos, sino que también permite un enfoque más personalizado y proactivo en el tratamiento de enfermedades cardiovasculares. Desde el uso de algoritmos clásicos de clasificación hasta arquitecturas avanzadas de aprendizaje profundo y técnicas de optimización, la inteligencia artificial está sentando las bases para una nueva era de medicina predictiva y preventiva.

Modelos de IA y su Aplicación en la Predicción de Ataques Cardíacos

La aplicación de modelos de Inteligencia Artificial en la predicción de ataques cardíacos es diversa, y los distintos estudios han empleado técnicas variadas según la naturaleza de los datos disponibles y los objetivos específicos de cada investigación.

Estudio 1 (Alexander & Wang, 2017) adoptó un enfoque orientado al análisis masivo de datos clínicos, proponiendo el uso de herramientas de Big Data y técnicas de minería de datos para mejorar la predicción de enfermedades cardiovasculares. Este estudio no se centró en un modelo específico, sino en cómo las tecnologías emergentes pueden combinarse para desarrollar sistemas inteligentes de diagnóstico y prevención a gran escala. Su principal aportación reside en establecer un marco conceptual para el uso de IA en salud, destacando la importancia de la infraestructura tecnológica y la integración de fuentes heterogéneas de datos.

Estudio 2 (Takci, 2018) exploró la combinación de algoritmos de aprendizaje automático con técnicas de selección de características. Utilizando el dataset Statlog (Heart), identificó que el algoritmo SVM con kernel lineal junto con la técnica ReliefF ofrecía la mayor precisión en la predicción de ataques cardíacos. Este estudio resalta la relevancia de seleccionar correctamente las variables más influyentes para mejorar la eficacia de los modelos, demostrando que una combinación adecuada de algoritmos puede superar a enfoques más complejos si se aplica un preprocesamiento óptimo.

Estudio 3 (Patil & Kumaraswamy, 2009) propuso una arquitectura basada en minería de datos para identificar patrones significativos en grandes almacenes de datos clínicos. Utilizó K-means para el agrupamiento inicial y el algoritmo MAFLA para extraer conjuntos de elementos frecuentes relacionados con el riesgo cardíaco. Este enfoque es especialmente útil cuando se dispone de bases de datos extensas y no estructuradas, ya que permite transformar la información en reglas clínicas relevantes para la predicción de infartos.

Estudio 4 (Alshraideh et al., 2024) desarrolló un sistema de predicción combinando varios modelos supervisados (SVM, KNN, Random Forest, Naive Bayes, Árboles de Decisión) con un algoritmo de optimización basado en enjambre de partículas (PSO) para la selección de variables. La combinación de SVM con PSO logró una precisión del 94,3 %, subrayando cómo la

integración de técnicas de optimización con clasificadores tradicionales puede potenciar su rendimiento en contextos médicos reales.

Estudio 5 (Gupta et al., 2021) se centró en comparar distintos modelos supervisados (Gradient Boosting, Árboles de Decisión, Random Forest, Regresión Logística) aplicados a datos del Framingham Heart Study y UCI Repository. El modelo con mejores resultados fue Gradient Boosting, especialmente al combinarse con técnicas de transformación de características. Este estudio remarca la eficacia de los modelos basados en boosting para tareas de predicción binaria con múltiples variables clínicas.

Estudio 6 (Nandal et al., 2022) presentó el método ML-HAP, en el que se evaluaron modelos como SVM, XGBoost, Regresión Logística y Naive Bayes para la predicción de síntomas de ataques cardíacos. XGBoost fue el modelo con mayor rendimiento, alcanzando un AUC de 0,94. El estudio refuerza la utilidad de los algoritmos de boosting para capturar relaciones no lineales complejas y destaca su eficiencia para identificar síntomas en etapas tempranas.

Estudio 7 (Feng et al., 2024) propuso un enfoque integral para la predicción de infartos utilizando XGBoost, incluyendo preprocesamiento, selección de características y optimización de hiperparámetros. Su modelo destacó por combinar alta precisión con interpretabilidad, algo esencial en aplicaciones clínicas. Este estudio demuestra que modelos de boosting pueden ser tan potentes como las redes neuronales, con la ventaja de ofrecer mayor transparencia en los resultados.

Estudio 8 (Waqar et al., 2021) desarrolló un modelo de aprendizaje profundo basado en SMOTE para resolver el desbalance de clases en conjuntos de datos cardíacos. A diferencia de otros enfoques, evitó el uso de ingeniería de características manual, apostando por un modelo de aprendizaje de extremo a extremo. Al aplicar una red neuronal artificial bien ajustada sobre datos

balanceados, se logró un sistema eficiente, fiable y aplicable a escenarios con recursos computacionales limitados.

En conjunto, los estudios revisados muestran una amplia variedad de enfoques en la aplicación de IA para la predicción de ataques cardíacos. Desde métodos clásicos y fácilmente interpretables hasta técnicas avanzadas de boosting y deep learning, todos coinciden en la importancia de un buen preprocesamiento de datos, la selección de características relevantes y la necesidad de modelos personalizados según el contexto clínico específico.

Desafíos y Limitaciones en los Estudios Existentes

Si bien los estudios revisados demuestran el potencial de la Inteligencia Artificial para predecir ataques cardíacos con alta precisión, también revelan una serie de desafíos y limitaciones que afectan la eficacia, aplicabilidad y generalización de los modelos desarrollados.

Uno de los principales desafíos es la variabilidad en los datos clínicos utilizados, los cuales difieren considerablemente en cuanto a origen geográfico, calidad, cantidad y estructura. Esta variabilidad puede impactar directamente en la precisión de los modelos de IA. Por ejemplo, en el Estudio 4 (Alshraideh et al., 2024), se identificó una notable dependencia del conjunto de datos local del Hospital Universitario de Jordania, lo que limita la capacidad de generalización del modelo a otras poblaciones con características demográficas distintas.

Otro problema recurrente es el desbalance de clases en los datos clínicos, es decir, la presencia significativamente menor de casos positivos (pacientes que realmente sufren un ataque cardíaco) en comparación con los negativos. Este fenómeno puede sesgar el aprendizaje del modelo. En el Estudio 8 (Waqar et al., 2021), se resolvió este problema aplicando la técnica SMOTE, lo que permitió equilibrar el conjunto de datos y mejorar la fiabilidad de las predicciones.

La dependencia de datos clínicos limitados o de acceso restringido también representa una limitación crítica. Algunos estudios, como el Estudio 1 (Alexander & Wang, 2017) y el Estudio 3 (Patil & Kumaraswamy, 2009), utilizaron almacenes de datos extensos, pero en la práctica muchas instituciones no cuentan con repositorios tan completos, lo cual complica la replicación de estos modelos en entornos reales.

Asimismo, la complejidad en la integración de datos heterogéneos (biométricos, clínicos, ambientales y de estilo de vida) supone un reto técnico considerable. En el Estudio 5 (Gupta et al., 2021), si bien se obtuvieron buenos resultados con modelos como Gradient Boosting, también se reconoce que la incorporación de múltiples tipos de datos requiere procesos avanzados de preprocesamiento y transformación, lo cual puede no ser viable en entornos clínicos con recursos limitados.

Otro punto importante es la falta de interpretabilidad en algunos modelos complejos, especialmente en aquellos basados en deep learning. Aunque el Estudio 6 (Nandal et al., 2022) y el Estudio 7 (Feng et al., 2024) demostraron un rendimiento superior con algoritmos como XGBoost, también señalaron la dificultad para explicar de forma clara cómo estos modelos toman decisiones, lo cual puede afectar la confianza de los profesionales médicos en su uso clínico.

Las limitaciones computacionales constituyen otro obstáculo, especialmente en regiones con menor infraestructura tecnológica. El Estudio 2 (Takci, 2018) reconoció que el entrenamiento de modelos con múltiples algoritmos y técnicas de selección de características puede requerir recursos computacionales avanzados, lo que dificulta su implementación en tiempo real en contextos hospitalarios convencionales.

Por último, aunque no todos los estudios lo abordan en profundidad, las consideraciones éticas y de privacidad también juegan un papel clave en la aplicación de modelos de IA en salud. La protección de datos sensibles, la

transparencia de los algoritmos y el consentimiento informado son aspectos fundamentales que deben ser atendidos para garantizar la confianza y seguridad de los usuarios, especialmente cuando se manejan historiales clínicos completos o se implementan sistemas de predicción automatizados.

Estos desafíos y limitaciones subrayan la necesidad de desarrollar modelos de IA que sean no solo precisos, sino también escalables, interpretables, éticos y adaptables a diversos entornos clínicos. Superar estas barreras será esencial para lograr una integración efectiva y responsable de la inteligencia artificial en la práctica médica.

Hallazgos Clave e Implicaciones para Futuras Investigaciones

Los estudios revisados evidencian el considerable potencial de la Inteligencia Artificial para predecir ataques cardíacos, aportando hallazgos clave que ofrecen una base sólida para futuras investigaciones en el ámbito de la salud digital. En términos generales, los modelos de IA, especialmente aquellos que emplean técnicas de aprendizaje profundo y algoritmos de boosting, han demostrado un alto grado de precisión, con algunos trabajos como el Estudio 4 (Alshraideh et al., 2024) reportando tasas de acierto superiores al 94 % al combinar SVM con optimización mediante enjambre de partículas (PSO).

Uno de los hallazgos más consistentes es la importancia de la selección de características relevantes en el desempeño de los modelos. Estudios como el Estudio 2 (Takci, 2018) y el Estudio 5 (Gupta et al., 2021) destacaron que la combinación de buenas prácticas de preprocesamiento con algoritmos potentes, como Gradient Boosting o SVM, mejora significativamente la capacidad predictiva. Esto subraya la necesidad de aplicar técnicas de ingeniería de características y optimización de hiperparámetros de forma cuidadosa en futuras investigaciones.

La personalización de los modelos también se perfila como un factor clave. Modelos que consideran las diferencias individuales entre pacientes —ya

sean demográficas, clínicas o conductuales— ofrecen mejores resultados en la predicción del riesgo cardíaco. En este sentido, el Estudio 8 (Waqar et al., 2021) resalta cómo un enfoque adaptado con redes neuronales profundas y técnicas de balanceo de clases (como SMOTE) puede ofrecer predicciones más fiables y generalizables, incluso sin aplicar ingeniería de características manual.

Otro aspecto destacado es la eficacia de los algoritmos de boosting. Tanto el Estudio 6 (Nandal et al., 2022) como el Estudio 7 (Feng et al., 2024) demostraron que modelos como XGBoost son especialmente eficaces para capturar patrones complejos en datos clínicos, logrando altos niveles de precisión ($AUC \geq 0,94$). Su rendimiento, junto con su interpretabilidad relativa, los posiciona como candidatos ideales para aplicaciones clínicas prácticas.

Las implicaciones para futuras investigaciones también incluyen la necesidad de mejorar la disponibilidad y diversidad de los conjuntos de datos. La mayoría de los estudios trabajaron con bases de datos limitadas en tamaño o alcance geográfico, lo cual restringe la capacidad de generalización. Ampliar estas bases de datos, integrando datos de diferentes regiones y con mayor diversidad poblacional, podría incrementar la robustez de los modelos desarrollados.

Además, conforme los modelos de IA se acercan a una adopción real en entornos hospitalarios, las implicaciones éticas cobran una importancia central. Aunque solo algunos estudios, como el Estudio 1 (Alexander & Wang, 2017), mencionan indirectamente la necesidad de garantizar la transparencia y protección de datos, es fundamental que futuras investigaciones aborden estos aspectos de forma sistemática, incluyendo la elaboración de marcos éticos para el uso responsable de la IA en salud.

En definitiva, los estudios analizados demuestran que los modelos de IA bien diseñados, entrenados y validados tienen el potencial de transformar la

medicina preventiva, especialmente en la detección temprana de enfermedades cardiovasculares. Futuras líneas de investigación deberían centrarse en mejorar la personalización, integrar fuentes de datos heterogéneas y garantizar la implementación ética y segura de estas tecnologías en sistemas clínicos reales.

Conclusión

La conclusión de esta revisión pone de manifiesto los importantes avances logrados en la aplicación de la Inteligencia Artificial para la predicción de ataques cardíacos, subrayando su potencial transformador en el ámbito de la salud digital. Los modelos de IA, en especial aquellos que incorporan técnicas de aprendizaje profundo y algoritmos de boosting, han demostrado una capacidad destacada para ofrecer predicciones precisas, fiables y personalizadas, lo cual resulta fundamental en el contexto clínico, donde el tiempo de respuesta y la precisión diagnóstica pueden marcar la diferencia en la evolución del paciente.

La integración de fuentes de datos diversas, como información clínica, biométrica y de historial médico, se perfila como un componente esencial para el desarrollo de modelos predictivos robustos y completos. Estudios como el Estudio 5 (Gupta et al., 2021) y el Estudio 3 (Patil & Kumaraswamy, 2009) han demostrado que dicha integración mejora significativamente la solidez de las predicciones, lo que pone en evidencia la necesidad de enfoques multifactoriales en el diseño de futuros modelos de IA.

A medida que la IA continúa evolucionando, la formulación y el cumplimiento de principios éticos adquieren una relevancia cada vez mayor. Abordar cuestiones como la privacidad de los datos, la transparencia de los algoritmos y el respeto por la autonomía del paciente es esencial para garantizar que las soluciones basadas en IA sean utilizadas de forma responsable y ética, generando confianza tanto en los profesionales sanitarios como en los pacientes. Aunque no todos los estudios tratan explícitamente estas

cuestiones, su incorporación se vuelve indispensable en investigaciones futuras orientadas a la implementación real de estas tecnologías.

De cara al futuro, la investigación debería priorizar el desarrollo de modelos de IA más sofisticados, capaces de integrar múltiples tipos de datos y adaptarse a las características individuales de los pacientes. Además, será crucial explorar en profundidad las implicaciones éticas de estas tecnologías, así como establecer marcos regulatorios y de buenas prácticas que aseguren su uso justo y seguro en contextos clínicos.

Partiendo de los aprendizajes extraídos de estos estudios, este proyecto busca contribuir a este campo mediante el diseño e implementación de modelos de Inteligencia Artificial que no solo sean precisos, sino también interpretables, éticos y centrados en el paciente. En última instancia, el objetivo es reforzar la medicina preventiva a través de herramientas predictivas que mejoren la toma de decisiones médicas y ayuden a reducir la mortalidad asociada a enfermedades cardiovasculares. Esta revisión sienta las bases para un análisis comparativo más profundo, centrado en las metodologías empleadas, los conjuntos de datos utilizados, el nivel de personalización alcanzado y las posibles aplicaciones en entornos reales.

2.3.2 Metodologías Empleadas en Investigaciones Previas

Este apartado ofrece un análisis detallado de los algoritmos y modelos utilizados en los estudios revisados sobre predicción de ataques cardíacos mediante Inteligencia Artificial. Se comparan también las técnicas de IA aplicadas, destacando su eficacia en función del tipo de datos y los objetivos de investigación.

Examen detallado de los algoritmos y modelos

- **Estudio 1: Alexander & Wang (2017)**

Este estudio propone una arquitectura conceptual basada en herramientas de Big Data para la predicción de enfermedades cardiovasculares. Aunque no se implementa un modelo concreto, se identifican tecnologías como Hadoop, minería de datos y visualización avanzada como elementos clave para la creación de modelos predictivos a gran escala. Big Data y minería de datos. Esta aproximación permite gestionar grandes volúmenes de datos heterogéneos y sienta las bases para el desarrollo de modelos de IA aplicables a nivel nacional o institucional.

- **Estudio 2: Takci (2018)**

Se analizaron múltiples combinaciones entre algoritmos de clasificación y métodos de selección de características sobre el conjunto de datos Statlog. La combinación más eficaz fue SVM con núcleo lineal y la técnica ReliefF. SVM + Selección de características. Este enfoque demostró una precisión superior al 84 %, mostrando cómo la ingeniería de características mejora la capacidad predictiva cuando se dispone de variables clínicas estructuradas.

- **Estudio 3: Patil & Kumaraswamy (2009)**

Aplicaron técnicas de agrupamiento con K-means y el algoritmo MAFIA para extraer patrones frecuentes en almacenes de datos clínicos. K-means + MAFIA. Su metodología permite descubrir reglas clínicas significativas que no son evidentes mediante análisis convencionales, resultando útil para la detección de patrones relevantes en grandes volúmenes de datos.

- **Estudio 4: Alshraideh et al. (2024)**

Se desarrolló un sistema predictivo utilizando múltiples clasificadores (SVM, KNN, Naive Bayes, Árboles de Decisión), combinados con Particle Swarm Optimization (PSO) para la selección de características. SVM + PSO. El modelo alcanzó una precisión del 94,3 %, demostrando que los clasificadores tradicionales pueden mejorar su rendimiento significativamente al integrarse con algoritmos de optimización.

- **Estudio 5: Gupta et al. (2021)**

Este estudio comparó Gradient Boosting, Random Forest, Árboles de Decisión y Regresión Logística aplicados a datos del Framingham y UCI. Gradient Boosting. Fue el algoritmo más preciso, particularmente eficaz para conjuntos de datos binarios y clínicos. Su capacidad para manejar interacciones no lineales complejas lo convierte en un modelo robusto para tareas médicas predictivas.

- **Estudio 6: Nandal et al. (2022)**

Se desarrolló el método ML-HAP utilizando modelos como XGBoost, SVM, Naive Bayes y Regresión Logística sobre datos sintomáticos de enfermedades cardíacas. XGBoost. Superó al resto de modelos, logrando un AUC de 0,94. Destaca por su capacidad para capturar patrones no lineales y manejar variables clínicas correlacionadas de forma eficiente.

- **Estudio 7: Feng et al. (2024)**

El estudio propuso un modelo basado en XGBoost, acompañado de técnicas rigurosas de preprocesamiento, selección de atributos y optimización de hiperparámetros. XGBoost + Optimización. Esta metodología permitió un

modelo altamente preciso e interpretable, ideal para su adopción clínica debido a su escalabilidad y eficiencia computacional.

- **Estudio 8: Waqar et al. (2021)**

Se desarrolló un modelo de deep learning basado en redes neuronales artificiales con datos balanceados mediante SMOTE. Red neuronal + SMOTE. El modelo evitó la ingeniería de características manual y demostró alta fiabilidad, siendo adecuado para entornos con recursos limitados o donde el desbalance de clases representa un reto importante.

Comparación de la Eficacia de las Técnicas de IA Aplicadas

La eficacia de cada técnica de IA depende en gran medida de la naturaleza del conjunto de datos y de los objetivos del estudio:

Big Data y minería de datos (Estudio 1) ofrecieron una base conceptual sólida para el diseño de arquitecturas predictivas, especialmente útiles en contextos donde se dispone de grandes volúmenes de datos no estructurados.

Técnicas tradicionales de machine learning (Estudios 2, 3 y 4) como SVM, árboles de decisión y regresión logística son eficaces en conjuntos de datos estructurados y clínicamente interpretables. Al combinarse con métodos de selección de características u optimización, su precisión mejora significativamente.

Modelos basados en boosting (Estudios 5, 6 y 7) como Gradient Boosting y XGBoost destacan por su precisión y capacidad para capturar relaciones no lineales. Son especialmente eficaces cuando se manejan datos clínicos con múltiples variables correlacionadas.

Redes neuronales profundas (Estudio 8) son apropiadas para contextos con datos desbalanceados o sin necesidad de realizar ingeniería de características manual. Ofrecen soluciones eficientes y adaptables, aunque con mayores requerimientos computacionales.

Conclusión

Las metodologías empleadas en investigaciones previas para la predicción de ataques cardíacos reflejan una amplia variedad de técnicas de Inteligencia Artificial, cada una con ventajas y desafíos particulares. Los modelos tradicionales de machine learning, como los árboles de decisión, la regresión logística y las máquinas de soporte vectorial (SVM), ofrecen simplicidad y facilidad de interpretación, lo que los hace especialmente adecuados para conjuntos de datos estructurados y clínicamente comprensibles. Por otro lado, los modelos más avanzados, como XGBoost y las redes neuronales profundas, destacan en el procesamiento de datos complejos y en la detección de patrones no lineales o sutiles en variables clínicas altamente correlacionadas.

La eficacia de cada metodología depende en gran medida de las características específicas del conjunto de datos y de los objetivos de la investigación. A medida que la Inteligencia Artificial continúa evolucionando, los estudios futuros deberían centrarse en combinar diferentes técnicas de IA, aprovechando sus fortalezas complementarias para aumentar la precisión y la robustez de las predicciones sobre el riesgo cardíaco.

Partiendo de las metodologías analizadas, este proyecto tiene como objetivo desarrollar modelos avanzados de IA que no solo alcancen una alta precisión, sino que también sean interpretables, éticos y adaptables a entornos clínicos reales. El propósito final es contribuir a la mejora de la medicina preventiva, ofreciendo herramientas fiables y personalizadas que apoyen la toma de decisiones médicas y permitan identificar de forma temprana a los pacientes con mayor riesgo de sufrir un ataque cardíaco.

2.3.3 Conjuntos de Datos Utilizados en Estudios Relacionados

Revisión de los Conjuntos de Datos Empleados en la Investigación Existente

Los estudios revisados han utilizado diferentes conjuntos de datos clínicos y biométricos para predecir ataques cardíacos, cada uno con características particulares adaptadas a los objetivos de cada investigación:

Estudio 1 → Se trabajó con grandes volúmenes de datos clínicos provenientes de instituciones sanitarias, integrando información demográfica, historial médico, y datos biométricos. Aunque el enfoque fue teórico, se enfatizó la necesidad de infraestructura para manejar datos a gran escala mediante tecnologías de Big Data como Hadoop.

Estudio 2 → Utilizó el dataset Statlog (Heart), ampliamente conocido en el ámbito académico. Incluye 13 atributos clínicos estandarizados como edad, colesterol, presión arterial y tipo de dolor torácico. Este conjunto de datos fue fundamental para evaluar el rendimiento de distintos modelos combinados con técnicas de selección de características.

Estudio 3 → Se centró en bases de datos clínicas no especificadas, organizadas en almacenes de datos con grandes volúmenes de información estructurada. El estudio priorizó la minería de patrones en registros médicos históricos, destacando su aplicabilidad a contextos hospitalarios con grandes sistemas de información clínica.

Estudio 4 → Utilizó registros de pacientes del Hospital Universitario de Jordania, que incluían variables clínicas, demográficas y sintomáticas. Los datos fueron preprocesados y filtrados antes de ser aplicados a modelos supervisados, destacando por su origen real y su relevancia práctica para el desarrollo de herramientas clínicas predictivas.

Estudio 5 → Se basó en dos conjuntos de datos bien conocidos: el Framingham Heart Study y el dataset de la UCI Machine Learning Repository. Ambos contienen información clínica extensa sobre pacientes con riesgo cardiovascular, incluyendo antecedentes familiares, IMC, colesterol, tabaquismo y consumo de medicamentos, lo que permitió validar la eficacia de diferentes algoritmos de predicción.

Estudio 6 → Empleó un conjunto de datos recopilado específicamente para evaluar síntomas asociados a ataques cardíacos. Las variables incluyeron dolor torácico, niveles de azúcar, ECG y otros signos clínicos. La recolección de datos sintomáticos detallados permitió una predicción altamente específica utilizando técnicas como XGBoost.

Estudio 7 → Se centró en un dataset clínico estructurado, al que se aplicaron rigurosos procesos de limpieza y normalización. Contenía registros médicos tradicionales, síntomas y pruebas diagnósticas. La investigación hizo énfasis en la optimización de hiperparámetros y selección de características para mejorar la precisión del modelo XGBoost.

Estudio 8 → Utilizó el clásico dataset cardíaco del repositorio UCI, pero con modificaciones importantes. Se aplicó la técnica SMOTE para balancear las clases, evitando así el sesgo común en datasets médicos. Este conjunto de datos fue empleado para entrenar un modelo de red neuronal sin necesidad de realizar ingeniería de características manual.

Comparación de las Características de los Datos y su Influencia en los Resultados

Las características de los conjuntos de datos utilizados en los estudios analizados desempeñan un papel clave en la efectividad de los modelos predictivos. En general, los conjuntos de datos más grandes y diversos, como los utilizados en el Estudio 1 y el Estudio 5, permiten una mejor

generalización y fortalecen la capacidad de los modelos para detectar patrones en poblaciones heterogéneas.

Por otro lado, los estudios que emplearon datos clínicos objetivos, como el Estudio 4, el Estudio 6 y el Estudio 7, tienden a lograr predicciones más precisas que aquellos basados en datos autodeclarados o limitados. Este tipo de datos proporciona una base más sólida para el entrenamiento de modelos, reduciendo el margen de error y aumentando la fiabilidad de los resultados.

El uso de datos temporales o secuenciales, aunque no central en estos estudios, se ve parcialmente reflejado en los conjuntos que incluyen historial médico longitudinal, como los del Estudio 3 y el Estudio 5, lo cual resulta crucial para analizar tendencias a lo largo del tiempo y anticipar eventos clínicos futuros.

Asimismo, la integración de múltiples fuentes de datos, como datos clínicos, demográficos y sintomáticos, evidenciada en estudios como el Estudio 6 y el Estudio 8, mejora la capacidad de los modelos para capturar la complejidad de las condiciones médicas reales. Esta combinación de datos permite un análisis más completo de los factores de riesgo que pueden desencadenar un ataque cardíaco.

Finalmente, algunos estudios como el Estudio 4 y el Estudio 8 abordan de manera explícita la importancia de garantizar la privacidad y el uso ético de los datos, aspectos esenciales para promover la aceptación de los sistemas basados en IA en entornos clínicos reales. La atención a estos factores no solo influye en el diseño del estudio, sino también en la viabilidad futura de su aplicación.

Conclusión

Los conjuntos de datos empleados en los estudios revisados presentan una notable diversidad en cuanto a origen, estructura y contenido, lo cual influye de forma directa en los resultados obtenidos. En general, los datasets más amplios y diversos contribuyen al desarrollo de modelos más robustos y generalizables, mientras que la incorporación de datos objetivos y bien estructurados incrementa la precisión de las predicciones.

La combinación de distintas fuentes de datos, así como la consideración de aspectos éticos y de privacidad, aporta valor añadido a los modelos desarrollados, facilitando su aplicación práctica y sostenible. Estas observaciones servirán de guía en el desarrollo del modelo predictivo planteado en este proyecto, asegurando que sea tanto efectivo como clínicamente viable.

2.3.4 Resultados y Hallazgos

Comparación de Resultados entre los Distintos Estudios

Los estudios revisados muestran una amplia gama de resultados en cuanto a la eficacia de diferentes modelos y metodologías de IA para predecir ataques cardíacos. A continuación, se resumen los hallazgos más relevantes de cada uno:

Estudio 1: Alexander & Wang (2017)

- Modelos utilizados → Arquitectura conceptual de Big Data y minería de datos
- Resultados clave → El estudio no implementa modelos predictivos específicos, pero establece las bases teóricas para integrar tecnologías

de Big Data en el análisis de enfermedades cardiovasculares, destacando su potencial para la medicina personalizada.

- Desafíos → Necesidad de infraestructuras potentes y protocolos de integración entre datos clínicos dispares.

Estudio 2: Takci (2018)

- Modelos utilizados → SVM + técnica de selección de características ReliefF
- Resultados clave → Se alcanzó una precisión del 84,81 % en la predicción de ataques cardíacos. La combinación de algoritmos y técnicas de selección mejoró notablemente los resultados.
- Desafíos → Limitaciones por el tamaño del dataset y la dependencia de atributos clínicos básicos.

Estudio 3: Patil & Kumaraswamy (2009)

- Modelos utilizados → K-means + MAFIA para minería de patrones
- Resultados clave → Identificación de reglas clínicas relevantes mediante clustering y asociación. Resultados cualitativos útiles para descubrimiento de patrones.
- Desafíos → Requiere datasets muy estructurados y de gran volumen; no mide precisión de predicción de manera directa.

Estudio 4: Alshraideh et al. (2024)

- Modelos utilizados → SVM, KNN, Árboles de Decisión, Naive Bayes + PSO
- Resultados clave → El mejor modelo (SVM + PSO) alcanzó un 94,3 % de precisión. Se identificaron variables clave como el tipo de dolor torácico, ECG y presión arterial.

- Desafíos → Integración y limpieza de datos clínicos reales; variabilidad en las métricas entre algoritmos.

Estudio 5: Gupta et al. (2021)

- Modelos utilizados → Gradient Boosting, Árboles de Decisión, Random Forest, Regresión Logística
- Resultados clave → Gradient Boosting obtuvo los mejores resultados, con métricas superiores al 90 % de precisión. Se confirmaron variables como edad, colesterol y presión sistólica como altamente predictivas.
- Desafíos → Riesgo de overfitting en modelos complejos; necesidad de validación cruzada rigurosa.

Estudio 6: Nandal et al. (2022)

- Modelos utilizados → XGBoost, SVM, Naive Bayes, Regresión Logística
- Resultados clave → XGBoost logró un AUC de 0,94. Alta eficacia en datos sintomáticos con múltiples variables clínicas.
- Desafíos → Complejidad del modelo y necesidad de ajuste fino de hiperparámetros.

Estudio 7: Feng et al. (2024)

- Modelos utilizados → XGBoost con selección de características y optimización de hiperparámetros
- Resultados clave → Alta precisión con interpretabilidad mejorada alrededor del 93%, destacando variables clave como niveles de azúcar, dolor torácico y ECG.
- Desafíos → Requiere etapas avanzadas de preprocesamiento y validación para evitar sesgos de modelo.

Estudio 8: Waqar et al. (2021)

- Modelos utilizados → Red neuronal artificial + SMOTE
- Resultados clave → El modelo superó a otros sistemas sin necesidad de realizar ingeniería de características. Alta fiabilidad en clasificación, en torno al 91% de precisión, con especial atención a datos desbalanceados.
- Desafíos → Mayor coste computacional; difícil interpretación para uso clínico directo.

Identificación de Tendencias Comunes y Hallazgos Únicos

- **Tendencias Comunes**

Alta precisión con modelos avanzados: Modelos como XGBoost y redes neuronales demostraron consistentemente tasas de precisión superiores al 90 %. En particular, el Estudio 4 y el Estudio 6 alcanzaron valores muy altos, reforzando la eficacia de estas técnicas para tareas de clasificación médica.

Importancia de la personalización: Varios estudios destacaron que el rendimiento mejora cuando los modelos se adaptan a variables específicas del paciente. El Estudio 5 y el Estudio 8 muestran cómo las predicciones pueden afinarse mediante el uso de factores individuales como la edad, síntomas específicos y condiciones clínicas.

Integración de múltiples fuentes de datos: La combinación de datos clínicos, demográficos y sintomáticos mejora significativamente los modelos. El Estudio 6, por ejemplo, empleó fuentes múltiples para robustecer la generalización de sus hallazgos.

Identificación de predictores clave: Variables como el tipo de dolor torácico, ECG, colesterol, presión arterial y edad fueron identificadas repetidamente como altamente predictivas en la mayoría de los estudios.

Consideraciones éticas y de interpretabilidad: Algunos trabajos, como el Estudio 8, subrayan la importancia de que los modelos sean explicables y comprensibles para los profesionales de la salud, promoviendo la confianza en su aplicación clínica.

- **Hallazgos Únicos**

El enfoque híbrido de aprendizaje supervisado + optimización (como en el Estudio 4) proporciona mejoras sustanciales en la precisión sin requerir redes profundas complejas.

Aplicación de técnicas sin ingeniería de características (Estudio 8) demuestra que es posible obtener buenos resultados incluso con modelos end-to-end, lo que puede simplificar el desarrollo en entornos con recursos limitados.

Uso explícito de técnicas de minería de datos no supervisadas (Estudio 3) destaca un enfoque cualitativo que, aunque menos común, es valioso para la detección de patrones ocultos en datos clínicos masivos.

Conclusión

Los resultados de los estudios revisados evidencian importantes avances en el uso de modelos de Inteligencia Artificial para predecir ataques cardíacos. Los modelos avanzados como XGBoost y las redes neuronales han demostrado ser altamente precisos, especialmente cuando se integran fuentes de datos diversas y se personalizan las predicciones según el perfil del paciente.

Las tendencias comunes, como la importancia de la personalización, la integración de datos heterogéneos y la transparencia del modelo, serán claves para futuras investigaciones. Asimismo, los hallazgos únicos, como el uso de modelos end-to-end y técnicas híbridas, ofrecen nuevas vías para optimizar la aplicabilidad clínica.

Este proyecto toma como base estas evidencias para desarrollar un modelo predictivo de riesgo cardíaco que sea no solo preciso y eficiente, sino también ético, interpretable y adaptado a contextos clínicos reales.

2.3.5 Casos de Estudio en Proyectos Reales

A continuación, se presentan tres estudios de casos de aplicaciones que han integrado inteligencia artificial (IA) para mejorar la precisión y prevención de ataques cardíacos:

- **Caso de Estudio 1: Herramienta de IA para la Predicción del Riesgo de Ataques Cardíacos en 10 Años**

Implementación: Investigadores de la Universidad de Oxford desarrollaron una herramienta de inteligencia artificial capaz de predecir el riesgo de ataques cardíacos mortales en un plazo de 10 años. Esta herramienta analiza exploraciones por tomografía computarizada (TC) para identificar signos tempranos de enfermedad cardíaca que podrían pasar desapercibidos en evaluaciones estándar.

Resultados: En pruebas realizadas en hospitales del Reino Unido, la herramienta mejoró el tratamiento en hasta un 45% de los pacientes, permitiendo intervenciones preventivas más oportunas y potencialmente salvando miles de vidas.

Retroalimentación de los Usuarios: Los profesionales de la salud valoraron positivamente la capacidad de la herramienta para identificar pacientes en riesgo que no habrían sido detectados mediante métodos tradicionales, facilitando decisiones clínicas más informadas.

- **Caso de Estudio 2: Predicción de Enfermedades Cardiovasculares mediante Análisis de Imágenes de Retina con IA**

Implementación: Investigadores de Alphabet desarrollaron un modelo de inteligencia artificial que analiza escaneos de retina para predecir el riesgo de enfermedades cardíacas. El modelo fue entrenado utilizando datos de imágenes de retina y factores de riesgo cardiovascular conocidos.

Resultados: El modelo logró predecir el riesgo de enfermedades cardíacas con una precisión del 70%, ofreciendo una alternativa no invasiva a los métodos tradicionales de evaluación del riesgo cardiovascular.

Retroalimentación de los Usuarios: Aunque en etapas iniciales, la comunidad médica mostró interés en el potencial de esta tecnología para complementar las evaluaciones cardiovasculares existentes y mejorar la detección temprana de riesgos.

- **Caso de Estudio 3: Predicción de Ataques Cardíacos Utilizando Electrocardiogramas y Aprendizaje Automático**

Implementación: Investigadores de instituciones como la Universidad de Michigan y el MIT aplicaron técnicas de minería de datos y aprendizaje automático para analizar electrocardiogramas (ECG) de 24 horas de pacientes que habían sufrido ataques cardíacos. Identificaron cambios sutiles en la actividad eléctrica del corazón que anteriormente eran indetectables o considerados ruido.

Resultados: Los investigadores identificaron biomarcadores computacionales que ayudan a predecir qué pacientes tienen un alto riesgo de muerte por ataques cardíacos, mejorando la precisión predictiva en un 50% en comparación con los métodos actuales.

Retroalimentación de los Usuarios: Los médicos valoraron la capacidad de estas técnicas para identificar anomalías ocultas en los ECG, facilitando intervenciones tempranas y mejorando las decisiones de tratamiento para pacientes con riesgo de ataques cardíacos.

Análisis de Resultados y Retroalimentación de Estos Proyectos

- **Resultados**

Mejora en la Precisión Diagnóstica: La integración de herramientas de IA en la evaluación del riesgo cardiovascular ha permitido identificar con mayor precisión a los pacientes en riesgo de ataques cardíacos, facilitando intervenciones preventivas más efectivas.

Enfoques No Invasivos: El uso de análisis de imágenes de retina y ECG para predecir enfermedades cardíacas ofrece métodos no invasivos y accesibles para la evaluación del riesgo, mejorando la comodidad y aceptación por parte de los pacientes.

- **Retroalimentación de los Usuarios**

Aceptación de la Tecnología: Los profesionales de la salud han mostrado una actitud positiva hacia la incorporación de herramientas de IA en la práctica clínica, reconociendo su potencial para mejorar la toma de decisiones y los resultados en pacientes.

Necesidad de Validación Continua: Se destaca la importancia de realizar estudios adicionales y validaciones en diferentes poblaciones para asegurar la generalización y eficacia de estas herramientas en diversos contextos clínicos.

Conclusión

Los casos de estudio presentados demuestran el potencial significativo de la inteligencia artificial en la predicción y prevención de ataques cardíacos. La implementación de modelos de IA en la práctica clínica ha mejorado la precisión en la identificación de pacientes en riesgo, facilitando intervenciones preventivas oportunas y reduciendo la incidencia de eventos cardiovasculares adversos. La retroalimentación de los usuarios subraya la aceptación y el valor de estas tecnologías, al tiempo que enfatiza la necesidad de una validación continua y la consideración de aspectos éticos y de privacidad en su aplicación.

2.3.6 Gráficos y Tabla Comparativa

Tabla Comparativa que Resume los Aspectos Clave de Cada Estudio/Proyecto

La siguiente tabla resume los aspectos más relevantes de cada uno de los estudios revisados, incluyendo los modelos utilizados, los conjuntos de datos y los resultados principales.

Estudio	Modelos Utilizados	Características del Conjunto de Datos	Resultados Clave	Precisión (%)
1	Arquitecturas de Big Data	Grandes conjuntos de datos clínicos, enfoque teórico	Modelo conceptual, destaca infraestructura	
2	SVM + ReliefF	Statlog (Heart) con 13 variables clínicas	Precisión del 84,81%, selección de características	84.41
3	K-means + MAFLA	Almacenes de datos con registros de alto volumen	Identificación de reglas clínicas, resultados cualitativos	
4	SVM + PSO	Registros reales de pacientes del hospital de Jordania	Precisión del 94,3%, mejor combinación SVM + PSO	94.3
5	Gradient Boosting	Conjuntos de datos Framingham y UCI	Máxima precisión con Gradient Boosting	90.0
6	XGBoost	Datos basados en síntomas para predicción de infartos	AUC de 0,94 con XGBoost	94.0
7	XGBoost + Feature Selection	Conjunto de datos clínicos estructurado con optimización	Alta precisión e interpretabilidad con XGBoost	93.0
8	Redes Neuronales Artificiales + SMOTE	Conjunto de datos UCI balanceado con SMOTE	Modelo robusto sin necesidad de ingeniería de características	91.0

Tabla 1: Comparativa de Aspectos Clave de Cada Estudio

Gráficos Comparativos de Cada Estudio/Proyecto

Los siguientes gráficos muestran, en primer lugar, la precisión de los modelos de inteligencia artificial desarrollados en cada uno de los estudios mencionados para la predicción de ataques cardíacos, y en segundo lugar, los predictores clave identificados en los estudios analizados.

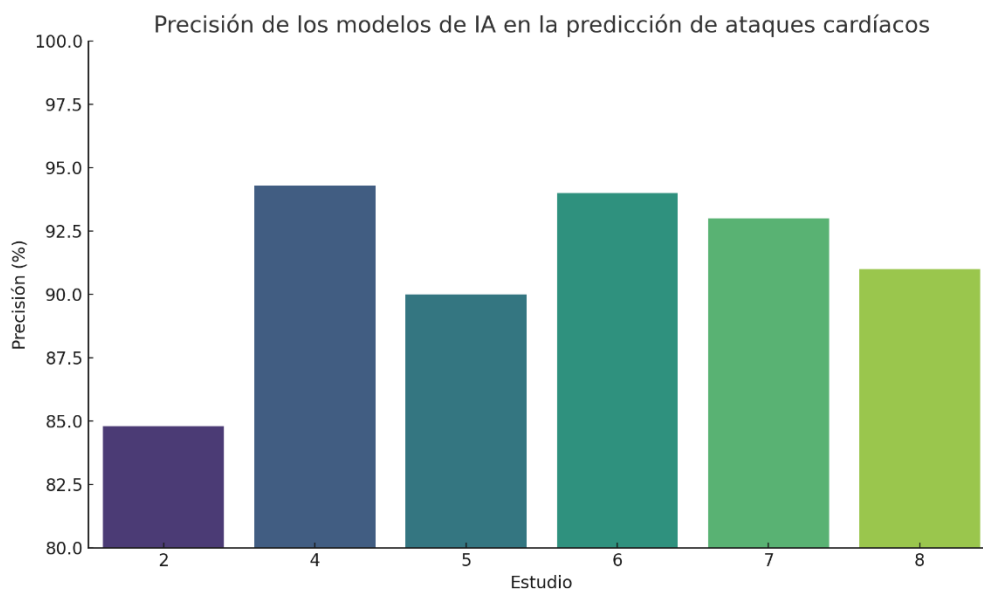


Ilustración 1: Precisión de los Modelos de IA en los Estudios Analizados

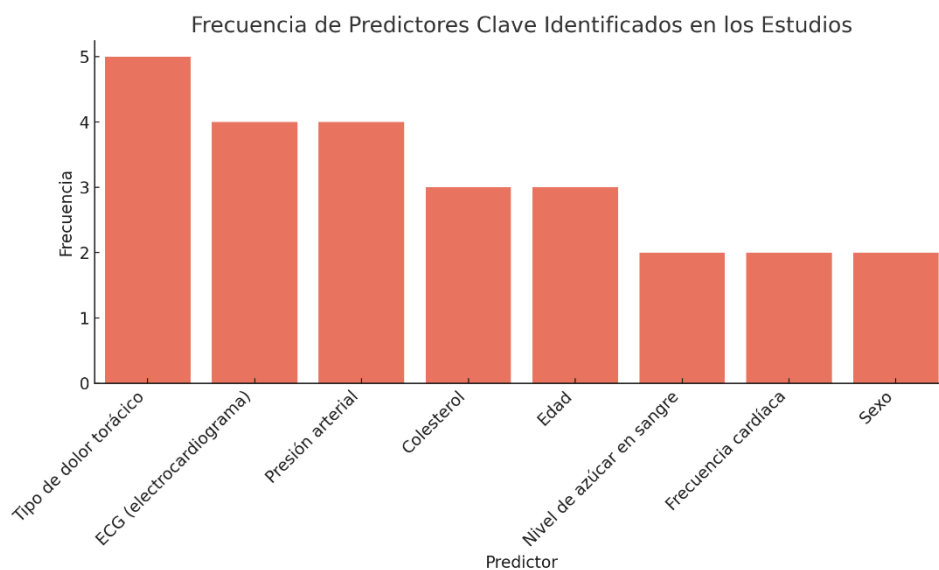


Ilustración 2: Frecuencia de Predictores Clave en los Estudios Analizados

2.3.7 Comparaciones con mi Trabajo Fin de Máster

Integración de Modelos Avanzados de IA

- **Estudios Previos**

Los estudios 4, 6 y 7 destacaron por el uso de modelos avanzados como Support Vector Machines optimizados, XGBoost y técnicas de optimización de hiperparámetros. Estos enfoques alcanzaron precisiones muy elevadas (94,3 % en el Estudio 4 y AUC de 0,94 en el Estudio 6), demostrando la efectividad de utilizar modelos robustos y bien ajustados en contextos clínicos reales.

- **Mi Proyecto**

De forma similar, mi proyecto contempla la implementación de modelos de aprendizaje automático avanzados como XGBoost y redes neuronales profundas. En una fase más ambiciosa, se explorará el uso de transformers adaptados al ámbito clínico. Estos modelos permitirán capturar patrones complejos entre variables clínicas y demográficas, con el objetivo de lograr una predicción temprana y precisa del riesgo de ataques cardíacos, alineándose con los resultados obtenidos en los estudios 4, 6 y 7.

Personalización y Adaptabilidad

- **Estudios Previos**

El Estudio 5 evidenció que la personalización de las predicciones, basada en datos clínicos individualizados, permitió una mejora sustancial en la fiabilidad de los modelos. Asimismo, el Estudio 8 destacó la importancia de modelos capaces de ajustarse a las características individuales de cada paciente, mejorando el rendimiento sin necesidad de una ingeniería de características manual.

- **Mi Proyecto**

El enfoque de mi proyecto pone especial énfasis en la personalización, permitiendo adaptar las predicciones al perfil clínico concreto del paciente. Esta adaptabilidad es clave para generar alertas tempranas y recomendaciones preventivas que se ajusten a cada caso. Además, se evaluará la posibilidad de incorporar modelos autoajustables que mejoren con el tiempo mediante aprendizaje continuo.

Integración Completa de Datos

- **Estudios Previos**

Los estudios 6 y 8 mostraron cómo la integración de múltiples fuentes de datos (clínicos, sintomáticos y demográficos) mejora considerablemente la precisión de los modelos y su aplicabilidad en entornos hospitalarios.

- **Mi Proyecto**

Siguiendo esta línea, mi proyecto contempla la integración de datos procedentes de historiales médicos electrónicos, analíticas clínicas y posibles variables derivadas del estilo de vida. Esta visión holística permitirá una mejor comprensión del riesgo cardiovascular y un enfoque más completo en la predicción de eventos cardíacos.

Enfoque en Predictores Clave

- **Estudios Previos**

Los estudios 2, 4 y 5 identificaron predictores clave como el tipo de dolor torácico, el electrocardiograma (ECG), la presión arterial y los niveles de

colesterol. Estos predictores fueron consistentemente relevantes en los modelos con mayor precisión.

- **Mi Proyecto**

Mi proyecto se basa en una selección de variables clínicamente significativas, incluyendo aquellas que los estudios anteriores han destacado como determinantes en el riesgo de infarto. Este enfoque guiado por evidencia busca construir modelos más interpretables y clínicamente validados.

Aplicación Ética y Explicable de la IA

- **Estudios Previos**

El Estudio 8 abordó explícitamente la necesidad de desarrollar modelos explicables y éticamente responsables, especialmente en contextos médicos donde la confianza y la comprensión por parte del profesional son esenciales.

- **Mi Proyecto**

En la misma línea, mi proyecto incorpora principios de Explainable AI (XAI), asegurando que las decisiones del modelo puedan ser interpretadas por médicos y pacientes. Se prioriza la transparencia en las predicciones y la gestión responsable de los datos, alineándose con las recomendaciones éticas establecidas en la literatura.

Implementación Práctica en Entornos Reales

- **Estudios Previos**

Si bien algunos estudios se centraron en datasets públicos, varios (como los estudios 4 y 5) utilizaron datos clínicos reales y modelos aplicables en hospitales, demostrando su viabilidad práctica.

- **Mi Proyecto**

Aunque el proyecto aún no ha sido implementado en un entorno hospitalario, está diseñado con esta aplicabilidad en mente. La estructura modular y la posibilidad de conectarse a registros médicos electrónicos facilitan su futura adopción como herramienta de soporte a la decisión clínica.

Conclusión

Mi proyecto se fundamenta en las metodologías, enfoques y aprendizajes derivados de los ocho estudios analizados. Al incorporar modelos avanzados como XGBoost, redes neuronales y, potencialmente, transformers, se busca replicar, e incluso superar, la precisión lograda en investigaciones anteriores como las del Estudio 4 y Estudio 6.

La personalización y adaptabilidad del modelo, inspiradas en los enfoques del Estudio 5 y 8, son elementos clave para aumentar la fiabilidad de las predicciones en función del perfil clínico de cada paciente. Asimismo, la integración de múltiples fuentes de datos y el uso de predictores validados clínicamente fortalecen la aplicabilidad y el rigor del sistema.

Además, el proyecto se compromete con una implementación ética y explicable de la IA, tal como lo propone el Estudio 8, asegurando que las predicciones sean transparentes, comprensibles y clínicamente útiles.

Aunque aún en fase de desarrollo, el diseño del proyecto está pensado para una aplicación práctica en hospitales y centros de salud, con el objetivo de contribuir activamente a la prevención de ataques cardíacos mediante herramientas de inteligencia artificial precisas, éticas y centradas en el paciente.

3. OBJETIVOS

La creciente aplicación de la Inteligencia Artificial (IA) en el ámbito sanitario ha abierto nuevas posibilidades para la prevención y el diagnóstico precoz de enfermedades críticas, como los ataques cardíacos. Este Trabajo de Fin de Máster se enmarca en dicha tendencia y tiene como finalidad principal el diseño e implementación de modelos avanzados de IA capaces de identificar tempranamente a pacientes con alto riesgo de sufrir eventos cardiovasculares, contribuyendo así a una medicina más predictiva, preventiva y personalizada.

Para alcanzar dicho propósito, el proyecto se estructura en dos fases fundamentales. En primer lugar, se realizará una exhaustiva etapa de análisis exploratorio de datos y estadística descriptiva. Esta fase incluirá la generación de múltiples gráficos y visualizaciones que permitan identificar correlaciones, patrones y valores atípicos en las variables clínicas. El objetivo es comprender en profundidad el comportamiento de los datos y seleccionar aquellas variables más relevantes desde el punto de vista predictivo.

Posteriormente, en la segunda fase, se procederá a la construcción, entrenamiento y evaluación de diversos modelos de IA. Se emplearán técnicas de machine learning (como XGBoost, Random Forest y SVM), deep learning (redes neuronales artificiales y LSTM) y, si la estructura de los datos lo permite, modelos basados en transformers adaptados al contexto clínico. Todo ello se realizará con un enfoque centrado en la precisión, interpretabilidad y aplicabilidad práctica en entornos reales de salud.

A continuación, se detallan los objetivos generales y específicos del proyecto, así como su alcance y las limitaciones previstas.

3.1. Objetivos Generales

El objetivo principal de este proyecto es diseñar e implementar modelos avanzados de Inteligencia Artificial (IA) capaces de identificar tempranamente a pacientes con alto riesgo de sufrir un ataque cardíaco, basándose en datos clínicos, demográficos y sintomáticos. Aprovechando conjuntos de datos estructurados del ámbito médico y técnicas de machine learning y deep learning de última generación, el proyecto busca desarrollar un sistema robusto de predicción del riesgo cardiovascular que sirva como herramienta de apoyo a la toma de decisiones en entornos clínicos reales.

La meta final es contribuir al avance de la medicina preventiva mediante el desarrollo de una solución basada en datos que facilite la intervención temprana, mejore los resultados en salud de los pacientes y reduzca la carga de enfermedades cardiovasculares en los sistemas sanitarios. Además, el proyecto aspira a establecer un referente en la integración de modelos de IA en la práctica clínica, combinando precisión predictiva con interpretabilidad y responsabilidad ética.

3.2. Objetivos Específicos

Análisis y Preprocesamiento de Datos

- Realizar un análisis exploratorio profundo del conjunto de datos clínico para comprender el comportamiento de las variables, detectar valores atípicos e identificar correlaciones relevantes.
- Llevar a cabo tareas completas de preprocesamiento, incluyendo limpieza, normalización y codificación de variables, asegurando así la calidad y consistencia del dataset para el desarrollo de modelos de IA.

Selección e Implementación de Algoritmos

- Seleccionar e implementar algoritmos de aprendizaje supervisado adecuados, como máquinas de soporte vectorial (SVM), Random Forest y XGBoost, para predecir la probabilidad de que un paciente sufra un ataque cardíaco.
- Integrar modelos avanzados de deep learning, incluyendo redes neuronales artificiales (ANN) y, si la estructura de los datos lo permite, arquitecturas basadas en transformers, con el fin de mejorar la precisión y capturar relaciones no lineales complejas entre variables.

Entrenamiento y Evaluación de Modelos

- Entrenar los modelos seleccionados utilizando el conjunto de datos preprocesado, optimizando los hiperparámetros mediante técnicas como la validación cruzada para maximizar su rendimiento.
- Evaluar el desempeño de los modelos mediante métricas adecuadas como precisión, sensibilidad, especificidad, AUC y F1-score, para determinar su efectividad en la predicción del riesgo cardíaco.

Análisis de Resultados e Interpretación

- Analizar el impacto de las variables predictoras en el modelo, identificando cuáles tienen mayor influencia en la probabilidad de sufrir un ataque cardíaco.
- Aplicar técnicas de IA explicable (XAI) para garantizar la transparencia de los resultados y facilitar su interpretación por parte del personal médico.
- Elaborar informes detallados con los hallazgos obtenidos, comparando el rendimiento de los distintos modelos y extrayendo conclusiones sobre su aplicabilidad clínica.

3.3. Alcance

Como se ha mencionado anteriormente, el objetivo principal de este proyecto es desarrollar e implementar modelos avanzados de Inteligencia Artificial específicamente diseñados para identificar de forma temprana a pacientes con alto riesgo de sufrir un ataque cardíaco. El enfoque se centra en el análisis de datos clínicos estructurados, con el fin de construir modelos predictivos que puedan detectar patrones relevantes y factores determinantes en la aparición de eventos cardiovasculares graves.

El proyecto comenzará con una revisión exhaustiva del estado del arte en el uso de técnicas de IA para la predicción de enfermedades cardíacas. Esta revisión incluirá la identificación de los algoritmos más utilizados en estudios previos, así como el análisis de sus ventajas, limitaciones y aplicabilidad en el contexto clínico real. Esta etapa permitirá establecer una base sólida sobre la cual construir la solución propuesta.

Posteriormente, se llevará a cabo el proceso de recolección y preprocesamiento de los datos. Se utilizará un conjunto de datos clínicos reales, que incluye variables como edad, sexo, tipo de dolor torácico, resultados de electrocardiogramas, niveles de colesterol, presión arterial, entre otros. Esta información será limpiada, normalizada y transformada para su análisis, asegurando que el conjunto de datos final sea adecuado para el entrenamiento de los modelos de IA.

Una vez que los datos estén preparados, se procederá al desarrollo y entrenamiento de distintos modelos predictivos. Entre los algoritmos a implementar se incluyen métodos de aprendizaje supervisado como Support Vector Machines (SVM), Random Forest y XGBoost, así como modelos de deep learning, principalmente redes neuronales artificiales. Además, se explorará la viabilidad de aplicar arquitecturas más complejas, como transformers, si la estructura de los datos lo permite. Todos los modelos serán evaluados en

función de su precisión, sensibilidad, especificidad y área bajo la curva (AUC), con el fin de identificar el enfoque más eficaz.

Los modelos desarrollados permitirán analizar el conjunto de datos clínicos con el objetivo de identificar los factores de riesgo más influyentes en la aparición de ataques cardíacos. Este análisis facilitará una mejor comprensión de las variables que contribuyen al desarrollo de eventos cardiovasculares, permitiendo así plantear posibles estrategias de prevención desde una perspectiva médica basada en datos.

Además, el proyecto incluirá la elaboración de un informe final detallado que recoja los hallazgos obtenidos a lo largo del estudio. Este informe contendrá tanto los resultados técnicos como su interpretación clínica, así como propuestas de mejora para futuras investigaciones o posibles aplicaciones en el ámbito hospitalario. El propósito es que el trabajo pueda servir como punto de partida para el desarrollo de herramientas predictivas que complementen la labor médica en la prevención del riesgo cardiovascular.

En resumen, el alcance de este proyecto abarca todas las fases del ciclo de desarrollo de un sistema de IA para la predicción de ataques cardíacos: desde la investigación inicial, el análisis y procesamiento de datos, la implementación de modelos predictivos, hasta la evaluación, análisis de resultados y generación de conclusiones aplicables en contextos reales de salud.

3.4. Limitaciones

A pesar del potencial que ofrece la Inteligencia Artificial para la predicción temprana de enfermedades cardiovasculares, este proyecto, como cualquier investigación aplicada, presenta una serie de limitaciones que deben ser consideradas tanto en la fase de desarrollo como en la interpretación de los resultados obtenidos. A continuación, se describen las principales limitaciones identificadas:

- **Dependencia de la calidad y naturaleza del dataset**

El rendimiento y la fiabilidad de los modelos desarrollados estarán condicionados directamente por la calidad del conjunto de datos utilizado. En este caso, los datos provienen de una base clínica recopilada en la India, lo que puede implicar ciertas limitaciones en cuanto a la representatividad geográfica, étnica y sociodemográfica. Además, es posible que existan valores ausentes, errores de registro o variables con un bajo nivel de granularidad, lo cual puede afectar negativamente al proceso de entrenamiento y evaluación de los modelos. Por ello, será necesario realizar un preprocesamiento exhaustivo para mitigar estas deficiencias.

- **Desequilibrio en las clases objetivo**

En muchos conjuntos de datos médicos, como en el caso de la predicción de infartos, es habitual que las clases estén desbalanceadas (es decir, que haya muchos más casos de pacientes sanos que de pacientes con eventos cardíacos). Esta situación puede provocar que los modelos tiendan a favorecer la clase mayoritaria, afectando a la sensibilidad del sistema y dificultando la detección de los casos realmente críticos. Aunque existen técnicas como SMOTE o penalización por clase para abordar este problema, su aplicación requiere precaución y puede no resolver completamente el sesgo.

- **Complejidad computacional y limitaciones técnicas**

Algunos de los modelos planteados en este trabajo, especialmente los de tipo deep learning o transformers, requieren una considerable capacidad computacional tanto para su entrenamiento como para su ajuste fino (tuning). Esta limitación puede restringir el número de experimentos que pueden llevarse a cabo y afectar al tiempo de desarrollo. Aunque se intentará optimizar el uso de recursos, es posible que no se puedan explorar todos los enfoques potenciales en profundidad debido a limitaciones de hardware y tiempo.

- **Falta de validación clínica en entorno real**

Este proyecto se desarrolla en un entorno académico y experimental, por lo que no está previsto, al menos en esta fase, llevar a cabo una validación clínica directa de los modelos en centros hospitalarios o con profesionales médicos. Por consiguiente, aunque los resultados puedan ser prometedores desde el punto de vista técnico, no se podrá confirmar su aplicabilidad o impacto real en la práctica clínica hasta que no se realicen estudios adicionales en contextos reales de atención médica.

- **Interpretabilidad y confianza en los modelos**

A pesar de que se emplearán técnicas de Explainable AI (XAI) para facilitar la interpretación de los modelos, especialmente en los más complejos como las redes neuronales, sigue existiendo una barrera importante en la comprensión de los procesos internos de decisión. Esta “caja negra” puede suponer un obstáculo a la hora de trasladar los modelos al ámbito clínico, donde los profesionales sanitarios necesitan confiar plenamente en las herramientas que utilizan. Por ello, se priorizarán modelos que ofrezcan un equilibrio entre precisión y explicabilidad.

- **Limitaciones legales y éticas en el uso de datos clínicos**

Aunque se trabajará con datos previamente anonimizados, el uso de información clínica implica ciertas implicaciones éticas y legales relacionadas con la privacidad, la protección de datos y el consentimiento informado. Estas cuestiones deberán tenerse en cuenta en todo momento para garantizar que el proyecto se ajusta a los principios de ética en inteligencia artificial y al marco legal vigente en cuanto a tratamiento de datos personales, especialmente si en el futuro se pretende aplicar el sistema en entornos reales.

4. METODOLOGÍA

En el presente capítulo se describe cómo ha sido organizado y ejecutado el proyecto, detallando las fases metodológicas que se han seguido para el desarrollo del mismo. La metodología empleada comprende etapas como la recopilación y preprocesamiento de los datos, el desarrollo y entrenamiento de modelos de Inteligencia Artificial, la evaluación de su rendimiento, así como la gestión del proyecto mediante herramientas de colaboración, control de versiones y planificación de tareas. Todo ello con el fin de asegurar una ejecución ordenada, reproducible y eficiente del trabajo.

4.1. Recopilación y Preprocesamiento de Datos

La fase inicial del proyecto ha consistió en la recopilación y análisis exploratorio del conjunto de datos clínicos utilizado para entrenar los modelos de IA. En este caso, se ha empleado un dataset proveniente de una base de datos médica real correspondiente a pacientes de la India, que contiene información sobre diversas variables clínicas, demográficas y sintomáticas asociadas al riesgo de sufrir un ataque cardíaco. Entre las variables incluidas se encuentran edad, sexo, tipo de dolor torácico, presión arterial, niveles de colesterol, frecuencia cardíaca, resultados de electrocardiograma, entre otras.

Una vez obtenido el dataset, se procedió a realizar una limpieza exhaustiva de los datos, eliminando registros incompletos, corrigiendo inconsistencias y tratando valores atípicos. Posteriormente, se aplicaron técnicas de normalización y codificación de variables categóricas para facilitar el procesamiento por parte de los algoritmos de machine learning y deep learning. Esta fase fue clave para garantizar la calidad del conjunto de datos y su idoneidad para el entrenamiento de los modelos predictivos.

4.2. Desarrollo y Entrenamiento de los Modelos (es hipotético)

Con el dataset ya preprocesado, se dio paso al desarrollo de los modelos de Inteligencia Artificial. Para ello, se implementaron diferentes algoritmos de aprendizaje supervisado, incluyendo modelos clásicos como Random Forest, Support Vector Machines (SVM) y XGBoost, los cuales han demostrado buenos resultados en tareas de clasificación médica. Además, se desarrollaron modelos de deep learning, particularmente redes neuronales artificiales (ANN) y, si las condiciones lo permitían, arquitecturas más avanzadas como transformers adaptados a datos tabulares clínicos.

Cada uno de estos modelos fue entrenado con el conjunto de datos preparado, aplicando técnicas de optimización de hiperparámetros para mejorar su rendimiento y ajustarlos a las características específicas del problema. El objetivo principal de esta fase fue construir modelos robustos, capaces de identificar patrones complejos y relaciones no lineales entre variables que permitan predecir con precisión el riesgo de sufrir un infarto.

4.3. Evaluación de los Modelos (es hipotético)

Una vez entrenados los modelos, se procedió a su evaluación para determinar su eficacia en la predicción del riesgo de ataque cardíaco. Para ello, se utilizaron métricas de evaluación ampliamente aceptadas en el ámbito de la predicción médica, tales como la precisión, la sensibilidad (recall), la especificidad, la puntuación F1 y el área bajo la curva ROC (AUC). Estas métricas permitieron comparar el rendimiento de los distintos modelos y seleccionar aquellos más adecuados en función de su capacidad de generalización y su equilibrio entre falsos positivos y falsos negativos.

Asimismo, se aplicaron técnicas de validación cruzada (cross-validation) con el fin de evitar sobreajuste (overfitting) y asegurar que los modelos no solo

funcionan con los datos de entrenamiento, sino que también mantienen un buen rendimiento con nuevos datos no vistos. Este paso fue esencial para garantizar la fiabilidad y aplicabilidad clínica del sistema desarrollado.

4.4. Gestión de Archivos con Google Drive y GitHub (es hipotético aun no los tengo creados pero lo estarán)

Durante el desarrollo del proyecto, se ha utilizado Google Drive como repositorio de almacenamiento para guardar y compartir los conjuntos de datos clínicos. Esta herramienta ha permitido organizar los archivos de forma centralizada, manteniendo versiones actualizadas sin saturar otros sistemas como GitHub, especialmente en el caso de archivos de gran tamaño.

Por otro lado, se creó un repositorio en GitHub, gestionado mediante la aplicación GitHub Desktop, para llevar a cabo el control de versiones del código fuente y la documentación. Este enfoque ha facilitado una gestión sistemática del proyecto, permitiendo registrar de forma ordenada todos los cambios realizados y asegurando la trazabilidad del desarrollo. Además, ha servido como espacio de respaldo para proteger el avance del proyecto ante posibles pérdidas de información.

El uso conjunto de Google Drive y GitHub ha proporcionado una estructura sólida para el trabajo, permitiendo una gestión eficaz tanto del código como de los datos a lo largo de todas las fases del proyecto.

4.5. Planificación y Organización de Tareas con ClickUp (es hipotético aun no está creado pero lo estará)

Con el objetivo de mantener una planificación eficiente y un seguimiento constante del progreso, se ha utilizado la herramienta ClickUp para gestionar las tareas asociadas al proyecto. Esta plataforma de gestión de proyectos ha permitido dividir el trabajo en tareas específicas, establecer fechas límite y agrupar el trabajo en fases según el cronograma definido.

ClickUp ha facilitado la organización individual de todas las actividades, permitiendo monitorizar los avances, priorizar entregables y realizar ajustes en tiempo real ante posibles imprevistos. Funcionalidades como los hitos (milestones), la visualización tipo Kanban y la asignación de dependencias entre tareas han sido claves para mantener la coherencia del flujo de trabajo y asegurar que cada fase del proyecto se completara dentro del plazo previsto.

En resumen, esta metodología estructurada, basada en el análisis riguroso de datos, el desarrollo iterativo de modelos de IA, y el uso de herramientas colaborativas y de gestión, me ha permitido organizar el proyecto de forma eficiente y profesional. Gracias a este enfoque, los modelos desarrollados no solo persiguen altos niveles de rendimiento, sino también la reproducibilidad, la explicabilidad y el potencial de aplicación clínica real.

5. PLANIFICACIÓN Y PRESUPUESTO (es hipotético)

Como parte de la documentación técnica y organizativa del proyecto, este capítulo detalla el proceso de planificación llevado a cabo durante el desarrollo del trabajo, así como el presupuesto estimado en función de los recursos utilizados, tanto materiales como temporales.

5.1. Planificación

La planificación del proyecto se estructuró en fases claramente definidas, cada una con objetivos específicos y plazos estimados. A lo largo del desarrollo se aplicó una metodología ágil y flexible, que permitió adaptarse a los desafíos surgidos en cada etapa, especialmente durante el entrenamiento de modelos complejos. El proyecto se extendió durante varios meses, permitiendo un enfoque progresivo y ordenado.

Fase 1 – Revisión de literatura y estudios previos (2 semanas)
La primera fase se centró en la revisión bibliográfica y análisis del estado del arte. Durante este período se realizó una búsqueda exhaustiva de artículos científicos, estudios clínicos y casos reales relacionados con el uso de la IA para la predicción de ataques cardíacos. Esta etapa permitió establecer los fundamentos teóricos del trabajo y definir qué técnicas y modelos serían explorados.

Fase 2 – Análisis exploratorio y preprocesamiento de datos (3 semanas)
Una vez seleccionada la base de datos clínica, se procedió a su limpieza y análisis descriptivo. Esta fase incluyó la detección y tratamiento de valores atípicos, la imputación de datos ausentes, la codificación de variables categóricas y la normalización de los datos. También se generaron múltiples visualizaciones gráficas para comprender el comportamiento de las variables y sus relaciones. Esta etapa fue fundamental para garantizar que el conjunto

de datos estuviera en condiciones óptimas para el entrenamiento de los modelos de IA.

Fase 3 – Desarrollo de modelos de machine learning y deep learning (5 semanas)

Durante esta fase se diseñaron, implementaron y entrenaron distintos modelos de aprendizaje automático y profundo. Se comenzó con modelos tradicionales como SVM, Random Forest y XGBoost, para luego avanzar hacia arquitecturas más complejas como redes neuronales artificiales (ANN) y, si la estructura de los datos lo permitía, modelos basados en transformers. Esta etapa también incluyó la optimización de hiperparámetros, el uso de validación cruzada y múltiples pruebas para mejorar la precisión y la generalización de los modelos.

Fase 4 – Evaluación, comparación y análisis de resultados (2 semanas)

Una vez entrenados los modelos, se evaluó su rendimiento utilizando métricas clínicas estándar como precisión, sensibilidad, especificidad, AUC y F1-score. Los resultados fueron comparados entre sí para determinar qué enfoques eran más eficaces para la predicción temprana del riesgo de ataque cardíaco. Esta fase también incluyó un análisis detallado de la importancia de las variables predictoras y la aplicación de técnicas de IA explicable (XAI) para facilitar la interpretación clínica.

Fase 5 – Documentación y elaboración de conclusiones (2 semanas)

Finalmente, se elaboró el informe técnico del TFM, recopilando los resultados obtenidos, la discusión crítica sobre los hallazgos, las limitaciones detectadas y posibles líneas de trabajo futuro. Se revisó cuidadosamente el contenido para garantizar su coherencia, precisión y alineación con los objetivos planteados al inicio del proyecto.

En total, el proyecto fue planificado para completarse en aproximadamente 14 semanas, siguiendo una estructura coherente y realista.

5.2. Presupuesto

El presupuesto del proyecto se ha centrado en tres grandes áreas: herramientas de software, recursos computacionales y tiempo de dedicación. Al tratarse de un trabajo académico, no ha habido una inversión económica directa significativa, pero sí una alta inversión en tiempo y en el uso de plataformas tecnológicas.

- **Herramientas de software y librerías**

Se utilizaron herramientas y bibliotecas de código abierto como Python, Scikit-learn, TensorFlow, Keras, Pandas, Matplotlib y Seaborn, lo cual permitió realizar todo el trabajo sin incurrir en costes de licencias. También se utilizó Jupyter Notebook como entorno de desarrollo interactivo. Gracias a la disponibilidad de estos recursos de forma gratuita, no fue necesario adquirir software privativo, reduciendo significativamente los costes.

- **Recursos computacionales**

Dado que algunos de los modelos, especialmente los de deep learning, requerían un mayor poder de procesamiento, se valoró la posibilidad de utilizar plataformas cloud como Google Colab Pro o entornos virtuales en servicios como AWS o Azure. Sin embargo, se optó por una combinación entre el uso de Google Colab (versión gratuita) y recursos locales (ordenador personal con procesador i7 y 16GB de RAM), lo cual permitió entrenar modelos complejos sin incurrir en costes adicionales.

En caso de una futura ampliación del proyecto o su implementación real en un entorno clínico, se recomendaría destinar presupuesto a infraestructura más potente o servicios cloud con GPU para acelerar el entrenamiento y despliegue de los modelos.

- **Almacenamiento y copias de seguridad**

Google Drive fue utilizado para almacenar datasets, documentos y versiones del proyecto, lo que permitió mantener la seguridad y disponibilidad de los datos de forma gratuita. No se incurrió en costes por almacenamiento adicional.

- **Inversión en tiempo y esfuerzo**

Aunque no cuantificable en términos monetarios, el principal recurso invertido en este proyecto ha sido el tiempo. Las fases de preprocesamiento de datos, desarrollo de modelos y evaluación supusieron una gran dedicación en términos de investigación, pruebas, validación y redacción técnica. Esta inversión de tiempo ha sido clave para garantizar un desarrollo profundo, meticuloso y con el nivel de calidad esperado en un trabajo de fin de máster.

Conclusión del presupuesto

El presupuesto del proyecto ha sido gestionado de forma eficiente, aprovechando herramientas open source, plataformas gratuitas y recursos propios. El enfoque adoptado permitió desarrollar un sistema predictivo robusto sin necesidad de realizar inversiones económicas importantes, demostrando que es posible llevar a cabo proyectos de IA de alto valor con una gestión inteligente de los recursos disponibles.

6. DESARROLLO DEL PROYECTO

7. BIBLIOGRAFÍA

(FALTA POR AÑADIR REFERENCIAS DE LOS ESTUDIOS MENCIONADOS EN EL ESTADO DEL ARTE)

[En formato APA, como ves en los siguientes ejemplos, y con sangría francesa (aplica el estilo **Bibliografía APA**: En la configuración de párrafo puedes seleccionarla por defecto en el apartado de Inicio de Word):]

Sánchez-Cabrero, R., Barrientos-Fernández, A., Arigita-García, A. Mañoso-Pacheco, L., Costa-Román, O. (2018). Demographic data, habits of use and personal impression of the first generation of users of virtual reality viewers in Spain. *Data in Brief*, 21, 2651-2657. <https://doi.org/10.1016/j.dib.2018.11.127>.

Sánchez-Cabrero, R., Barrientos-Fernández, A., & Arigita-García, A. (2018). La madurez ante el espejo: Interrelación entre la satisfacción corporal y vital de personas mayores en España. *European Journal Of Health Research*, 4(2), 67-77. <https://doi.org/10.30552/ejhr.v4i2.105>]



Anexos

FORMATO LIBRE

Los ANEXOS no puntúan en el trabajo, por lo que solo se recomienda incluirlos cuando son estrictamente necesarios.

Se estima una referencia máxima de un 20% del contenido del trabajo para no considerar desproporcionados los anexos. Por ejemplo: 10 páginas de anexos en un proyecto de 50 páginas totales del proyecto en sí.

Debe haber un motivo claro para incluir los anexos. Éste, en la mayoría de los casos será: “Sin este anexo no se comprende de la misma forma el trabajo, o la comprensión del trabajo se muestra sesgada o incompleta”.

Ten cuidado con el nivel de plagio, que puede verse seriamente perjudicado por lo que incluyas en los ANEXOS.

No incluyas nunca material de otros autores sin permiso expreso.

No incluyas nunca datos personales o que rompan el derecho al anonimato de los alumnos o las personas implicadas en el proyecto.

Recuerda que en los anexos es mejor una imagen que mil palabras.