

TESIS TESIS TESIS TESIS TESIS

**UNIVERSIDAD AUTÓNOMA DE AGUASCALIENTES**

---

**Centro de Ciencias Básicas**

**Departamento de Sistemas de Información**



**UNIVERSIDAD AUTÓNOMA  
DE AGUASCALIENTES**

TESIS

**“Identificación de Factores de Riesgo de Cardiopatías  
Isquémicas en Mujeres Mediante Aprendizaje Automático”**

**PRESENTA:**

Ricardo Daniel Lozano Sanchez

**PARA OPTAR POR EL GRADO DE:**

*“Maestro en Informática y Tecnologías Computacionales”*

**TUTORAS:**

Dra. María Dolores Torres Soto

Dra. Aurora Torres Soto

**COMITÉ TUTORAL:**

Dra. Esperanza Sanchez Alemán

Aguascalientes, Ags. 27 de Agosto del 2024.

TESIS TESIS TESIS TESIS TESIS



**MTRO. EN C. JORGE MARTÍN ALFÉREZ CHÁVEZ**  
DECANO DEL CENTRO DE CIENCIAS BÁSICAS

PRESENTE

Por medio del presente como *TUTORA* designada del estudiante **Ricardo Daniel Lozano Sanchez** con ID **236294** quien realizó la tesis titulada: **"IDENTIFICACIÓN DE FACTORES DE RIESGO DE CARDIOPATÍAS ISQUÉMICAS EN MUJERES MEDIANTE APRENDIZAJE AUTOMÁTICO"**, un trabajo propio, innovador, relevante e inédito y con fundamento en el Artículo 175, Apartado II del Reglamento General de Docencia doy mi consentimiento de que la versión final del documento ha sido revisada y las correcciones se han incorporado apropiadamente, por lo que me permito emitir el **VOTO APROBATORIO**, para que él pueda proceder a imprimirla así como continuar con el procedimiento administrativo para la obtención del grado.

Pongo lo anterior a su digna consideración y sin otro particular por el momento, me permito enviarle un cordial saludo.

**ATENTAMENTE**

**"Se Lumen Proferre"**

Aguascalientes, Ags., a 22 de Julio de 2024.



**DRA. ESPERANZA SANCHEZ ALEMAN**  
Comité tutorial

c.c.p.- Interesado

c.c.p.- Secretaría Técnica del Programa de Posgrado

**MTRO. EN C. JORGE MARTÍN ALFÉREZ CHÁVEZ**  
DECANO DEL CENTRO DE CIENCIAS BÁSICAS

PRESENTE

Por medio del presente como *TUTORA* designada del estudiante **Ricardo Daniel Lozano Sanchez** con ID **236294** quien realizó la tesis titulada: **"IDENTIFICACIÓN DE FACTORES DE RIESGO DE CARDIOPATÍAS ISQUÉMICAS EN MUJERES MEDIANTE APRENDIZAJE AUTOMÁTICO"**, un trabajo propio, innovador, relevante e inédito y con fundamento en el Artículo 175, Apartado II del Reglamento General de Docencia doy mi consentimiento de que la versión final del documento ha sido revisada y las correcciones se han incorporado apropiadamente, por lo que me permito emitir el **VOTO APROBATORIO**, para que él pueda proceder a imprimirla así como continuar con el procedimiento administrativo para la obtención del grado.

Pongo lo anterior a su digna consideración y sin otro particular por el momento, me permito enviarle un cordial saludo.

**ATENTAMENTE**  
**"Se Lumen Proferre"**  
Aguascalientes, Ags., a 22 de Julio de 2024.



**DRA. AURORA TORRES SOTO**  
Co-Tutora de tesis

c.c.p.- Interesado  
c.c.p.- Secretaría Técnica del Programa de Posgrado

**MTRO. EN C. JORGE MARTÍN ALFÉREZ CHÁVEZ**  
DECANO DEL CENTRO DE CIENCIAS BÁSICAS

PRESENTE

Por medio del presente como *TUTORA* designada del estudiante **Ricardo Daniel Lozano Sanchez** con ID **236294** quien realizó la tesis titulada: **"IDENTIFICACIÓN DE FACTORES DE RIESGO DE CARDIOPATÍAS ISQUÉMICAS EN MUJERES MEDIANTE APRENDIZAJE AUTOMÁTICO"**, un trabajo propio, innovador, relevante e inédito y con fundamento en el Artículo 175, Apartado II del Reglamento General de Docencia doy mi consentimiento de que la versión final del documento ha sido revisada y las correcciones se han incorporado apropiadamente, por lo que me permito emitir el **VOTO APROBATORIO**, para que él pueda proceder a imprimirla así como continuar con el procedimiento administrativo para la obtención del grado.

Pongo lo anterior a su digna consideración y sin otro particular por el momento, me permito enviarle un cordial saludo.

**ATENTAMENTE**  
**"Se Lumen Proferre"**  
Aguascalientes, Ags., a 22 de Julio de 2024.

  
**DRA. MARIA DOLORES TORRES SOTO**  
Tutora de tesis

c.c.p.- Interesado  
c.c.p.- Secretaría Técnica del Programa de Posgrado

Fecha de dictaminación dd/mm/aaaa: 24/07/24

**NOMBRE:** Ricardo Daniel Lozano Sánchez **ID** 236294

**PROGRAMA:** Maestría en Informática y Tecnologías Computacionales **LGAC (del posgrado):** Ingeniería de sistemas decisionales para mejorar de procesos organizacionales

**TIPO DE TRABAJO:** ( x ) Tesis ( ) Trabajo Práctico

**TITULO:** Identificación de Factores de Riesgo de Cardiopatías Isquémicas en Mujeres Mediante Aprendizaje Automático

**IMPACTO SOCIAL (señalar el impacto logrado):**

Se trata de un proyecto en el que se estudia una de las patologías que provoca la mayor tasa de mortalidad a niveles mundial y nacional en mujeres (infarto de miocardio). El trabajo propone apoyar el diagnóstico temprano y efectivo en mujeres a través de mecanismos automatizados de inteligencia artificial. Este trabajo contó con la colaboración y orientación de especialistas en cardiología del Centenario Hospital Miguel Hidalgo (CHMH), que apoyaron en la validación de las herramientas desarrolladas.

**INDICAR SI NO N.A. (NO APLICA) SEGÚN CORRESPONDA:**

INDICAR	SI	NO	N.A. (NO APLICA)	SEGÚN CORRESPONDA:
<i>Elementos para la revisión académica del trabajo de tesis o trabajo práctico:</i>				
SI				El trabajo es congruente con las LGAC del programa de posgrado
SI				La problemática fue abordada desde un enfoque multidisciplinario
SI				Existe coherencia, continuidad y orden lógico del tema central con cada apartado
SI				Los resultados del trabajo dan respuesta a las preguntas de investigación o a la problemática que aborda
SI				Los resultados presentados en el trabajo son de gran relevancia científica, tecnológica o profesional según el área
SI				El trabajo demuestra más de una aportación original al conocimiento de su área
SI				Las aportaciones responden a los problemas prioritarios del país
SI				Generó transferencia del conocimiento o tecnológica
SI				Cumple con la ética para la investigación (reporte de la herramienta antiplagio)
<i>El egresado cumple con lo siguiente:</i>				
SI				Cumple con lo señalado por el Reglamento General de Docencia
SI				Cumple con los requisitos señalados en el plan de estudios (créditos curriculares, optativos, actividades complementarias, estancia, predoctoral, etc)
SI				Cuenta con los votos aprobatorios del comité tutoral, en caso de los posgrados profesionales si tiene solo tutor podrá liberar solo el tutor
SI				Cuenta con la carta de satisfacción del Usuario
SI				Coincide con el título y objetivo registrado
SI				Tiene congruencia con cuerpos académicos
SI				Tiene el CVU del Conacyt actualizado
NA				Tiene el artículo aceptado o publicado y cumple con los requisitos institucionales (en caso que proceda)
<i>En caso de Tesis por artículos científicos publicados</i>				
NA				Aceptación o Publicación de los artículos según el nivel del programa
NA				El estudiante es el primer autor
NA				El autor de correspondencia es el Tutor del Núcleo Académico Básico
NA				En los artículos se ven reflejados los objetivos de la tesis, ya que son producto de este trabajo de investigación.
NA				Los artículos integran los capítulos de la tesis y se presentan en el idioma en que fueron publicados
NA				La aceptación o publicación de los artículos en revistas indexadas de alto impacto

Con base a estos criterios, se autoriza se continúen con los trámites de titulación y programación del examen de grado:

SI  X  
No

**FIRMAS**

**Elaboró:**

\* NOMBRE Y FIRMA DEL CONSEJERO SEGÚN LA LGAC DE ADSCRIPCIÓN:

Dra. María Dolores Torres Soto

NOMBRE Y FIRMA DEL SECRETARIO TÉCNICO:

Dra. Lizeth Itziguery Solano Romo

\* En caso de conflicto de intereses, firmará un revisor miembro del NAB de la LGAC correspondiente distinto al tutor o miembro del comité tutoral, asignado por el Decano

**Revisó:**

NOMBRE Y FIRMA DEL SECRETARIO DE INVESTIGACIÓN Y POSGRADO:

Alejandro Padilla Díaz

**Autorizó:**

NOMBRE Y FIRMA DEL DECANO:

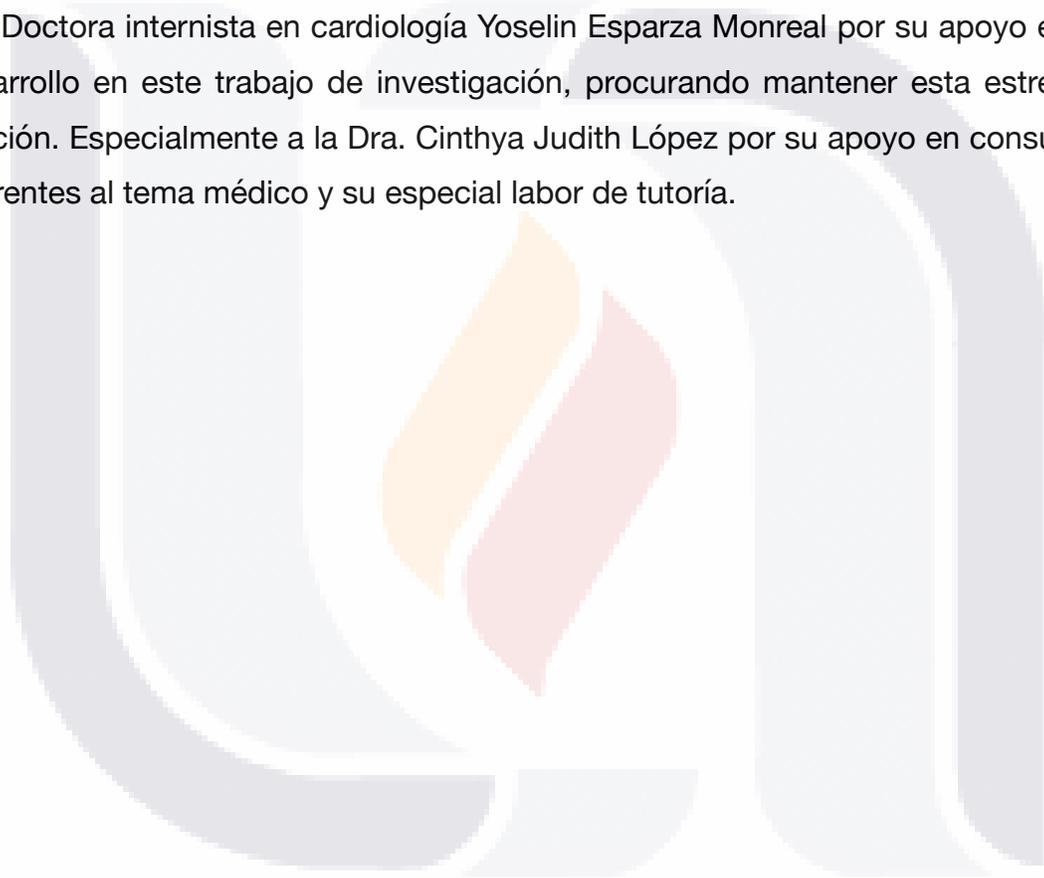
Mtro. En C. Jorge Martín Alférez Chávez

**Nota: procede el trámite para el Depto. de Apoyo al Posgrado**

En cumplimiento con el Art. 105C del Reglamento General de Docencia que a la letra señala entre las funciones del Consejo Académico: ... Cuidar la eficiencia terminal del programa de posgrado y el Art. 105F las funciones del Secretario Técnico, llevar el seguimiento de los alumnos.

## Agradecimientos

Al Honorable Centenario Hospital Hidalgo, al director de enseñanza Felipe de Jesús Flores Parkman, a la Dra. Cinthya Judith López, Directora de cardiología y a la Doctora internista en cardiología Yoselin Esparza Monreal por su apoyo en el desarrollo en este trabajo de investigación, procurando mantener esta estrecha relación. Especialmente a la Dra. Cinthya Judith López por su apoyo en consultas referentes al tema médico y su especial labor de tutoría.



## Dedicatorias

A la Universidad y al personal administrativo por otorgarme la oportunidad de continuar con mis estudios y poder hacer uso de las instalaciones para continuar con mi crecimiento profesional.

A mis tutoras, que gracias a su disposición adquirí las herramientas necesarias que me impulsaron a seguir con la investigación y aumentar mi conocimiento en el áreas de la inteligencia artificial en un contexto médico.

A mi familia, mis primos, y tíos: por su paciencia y apoyo incondicional. A mi abuelo; que me aportó herramientas para la vida, y especialmente a mi mamá, por acompañarme en este proceso tan importante de mi formación.

## Indice General

I. Introducción.....	7
II. Planteamiento del Problema.....	9
III. Justificación.....	11
Efectos en la comunidad .....	12
IV. Objetivos.....	16
V. Marco Contextual .....	17
VI. Marco Teórico .....	19
Sistema cardiovascular.....	19
Enfermedades del corazón .....	21
Factores de riesgo de enfermedad cardiovascular.....	22
Factores de riesgo en mujeres.....	25
Inteligencia artificial.....	27
VII. Metodología.....	32
Material .....	37
VIII.Resultados.....	43
Generación de modelos.....	44
Conjunto de datos de Framingham .....	44
Conjunto de datos combinado .....	45
Conjunto de datos CDC 2020.....	46
Conjunto de datos CDC 2022.....	47
Conjunto de datos Punjab, Pakistan .....	48
Servidor.....	49
Cuestionario WEB.....	51
Aplicación iOS.....	56
IX. Discusión de Resultados.....	61
Recomendaciones .....	62
X. Conclusiones .....	63
Trabajo Futuro.....	64
Bibliografía.....	65
Anexos.....	70
A. Certificados y Constancias .....	71
B. Artículo .....	76

**Indice de Tablas**

TABLA 1. Principales causas de mortalidad en el estado de Aguascalientes 2021. ....14

TABLA 2. Factores de riesgo modificables y no modificables. ....22

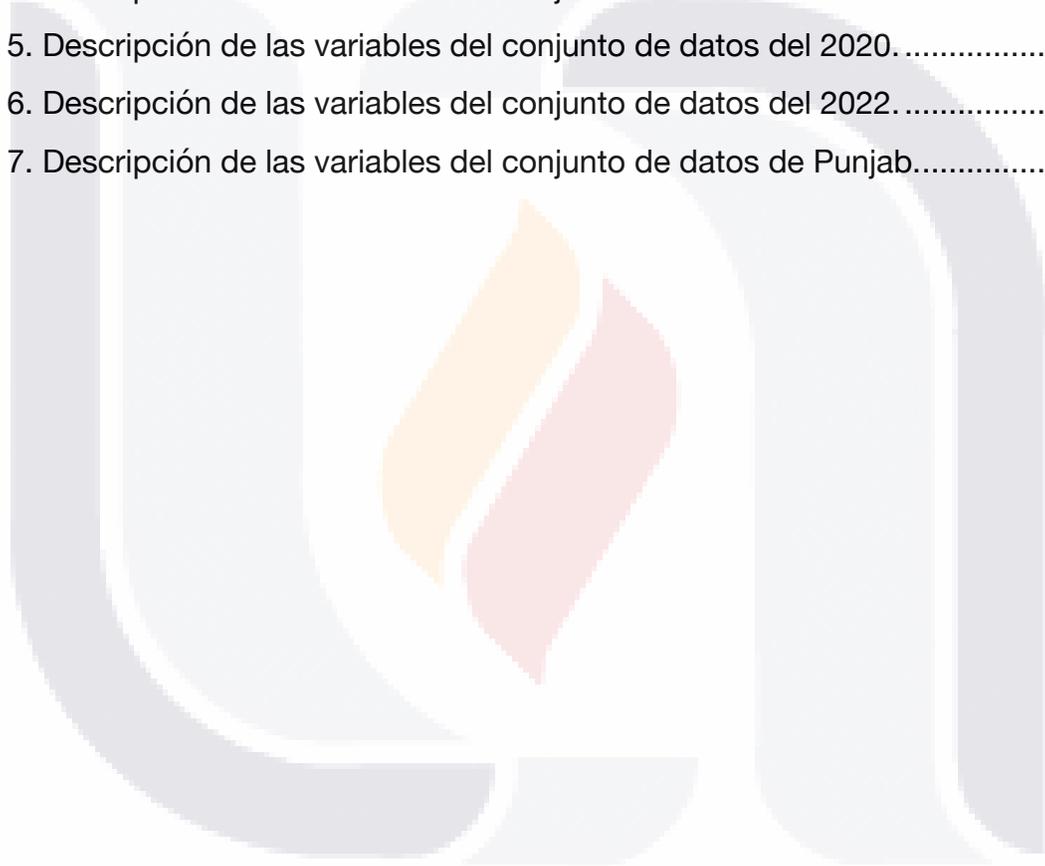
TABLA 3. Descripción de las variables del conjunto de datos de Framingham.....37

TABLA 4. Descripción de las variables del conjunto de datos combinados. ....38

TABLA 5. Descripción de las variables del conjunto de datos del 2020. ....39

TABLA 6. Descripción de las variables del conjunto de datos del 2022. ....40

TABLA 7. Descripción de las variables del conjunto de datos de Punjab.....42



## Indice de Figuras

FIG 1. Mortalidad por ECV para hombres y mujeres en Estados Unidos de 1979 a 2011 .....	11
FIG 2. Tasa de mortalidad de ECV por cada 100,000 habitantes por sexo en Mexico.....	12
FIG 3. Muertes para las mujeres y grupo de edad en Mexico por ECV. ....	13
FIG 4. Diagrama del corazón humano. ....	19
FIG 5. Diagrama del sistema eléctrico del corazón.....	20
FIG 6. Edad promedio y el periodo de factores de riesgo tradicionales y emergentes.....	25
FIG 7. Factores de riesgo tradicionales y emergentes en mujeres.....	26
FIG 8. Metodología: Identificación de factores de riesgo de ECV en la mujer.....	32
FIG 9. Metodología: Generación de modelos.....	34
FIG 10. Metodología: Cuestionarios de posible enfermedad cardiovascular. ....	36
FIG 11. Matriz de confusión, modelo 1.....	44
FIG 12. Matriz de confusión, modelo 2.....	45
FIG 13. Matriz de confusión, modelo 3.....	46
FIG 14. Matriz de confusión, modelo 4.....	47
FIG 15. Matriz de confusión, modelo 5.....	48
FIG 16. Captura de pantalla del proyecto del servidor en VS.....	49
FIG 17. Cuestionario del riesgo de ECV en 10 años. ....	51
FIG 18. Cuestionario de ECV de poblaciones variadas. ....	52
FIG 19. Cuestionario con datos de ECV del CDC 2020.....	53
FIG 20. Cuestionario con datos de ECV del CDC 2022.....	54
FIG 21. Cuestionario de mortalidad de IAM. ....	55
FIG 22. Cuestionario iOS: ECV en 10 años iOS. ....	56
FIG 23. Cuestionario iOS: ECV de poblaciones variadas. ....	57
FIG 24. Cuestionario iOS: ECV del CDC 2020. ....	58
FIG 25. Cuestionario iOS: ECV del CDC 2022. ....	59
FIG 26. Cuestionario iOS: mortalidad de IAM. ....	60

## Acrónimos

- EAC: Enfermedad de las arterias coronarias.
- IM: Infarto de miocardio.
- ECV: Enfermedades cardiovasculares.
- SICA: Síndrome isquémico coronario agudo.
- IAM: Infarto agudo de miocardio.
- IMC: Índice de masa corporal.
- SVM: Support Vector Machine.
- CDC: Centers for Disease Control and Prevention.
- BP: Presión arterial.
- DM: Diabetes mellitus.
- HDL: Lipoproteína de alta densidad.
- HTN: Hipertensión.
- IHD: Enfermedad cardíaca isquémica.
- LDL: Lipoproteína de baja densidad.
- TG: Triglicéridos.
- AUC: Área bajo la curva.

## Resumen

Los síntomas previos al infarto difieren entre hombres y mujeres, siendo menos notorios en el sexo femenino; esto es debido a diferencias fisiológicas y hormonales entre ambos sexos, siendo más sutiles y no claramente reconocibles por ellas o el personal médico, esto es debido a que históricamente han sido poco estudiados en mujeres. En el presente documento se explora la posibilidad de concebir una herramienta digital que ayude a identificar infarto en mujeres. Para conseguir esto proponemos hacer análisis de datos de mujeres que sufrieron un infarto o tienen afecciones coronarias similares, esto con el objetivo de lograr identificar factores de riesgo relacionados con este padecimiento que sean específicas para la mujer, y consecuentemente, obtener una herramienta clasificadora. En nuestros resultados obtuvimos que los factores de riesgo que mayor impacto generan en mujeres según las diversas fuentes utilizadas en los modelos de aprendizaje automático son: último grado de estudios, si es fumador actual, cuantos cigarrillos al día consume, si sufrió de un evento cerebrovascular, si sufre de hipertensión (incluyendo hipertensión medicada), su índice de masa corporal, su edad, el tipo de dolor de pecho, los niveles de glucosa en sangre, si tuvo diabetes gestacional, si padece enfermedad de los riñones, piezas dentales removidas, horas de sueño, grupo étnico, altura en metros, padecer VIH, padecer cáncer de piel, sufrir de alcoholismo, haber sido vacunada de neumonía y de la gripe. Nuestros distintos modelos obtuvieron una precisión de 86%, 100%, 70%, 86% y del 100%. Destacamos la importancia de incorporar en futuros estudios una mayor cantidad de factores de riesgo en mujeres, lo cual permitirá incrementar la precisión de modelos específicos en la clasificación de mujeres en riesgo de enfermedad cardiovascular y permitirá conocer mejor aún el impacto de cada uno de los factores de riesgo específicos en la mujer.

Palabras clave: Aprendizaje Automático, Datos clínicos, Sistema para el Apoyo en la Toma de Decisiones, Enfermedad Cardiovascular en Mujeres.

## Abstract

The symptoms prior to the infarction differ between men and women, being less noticeable in the female sex; this is due to physiological and hormonal differences between both sexes, being more subtle and not clearly recognizable by them or the medical personnel, this is due to the fact that historically they have been little studied in women. This document explores the possibility of conceiving a digital tool that helps identify heart attack in women. To achieve this, we propose to analyze data from women who suffered a heart attack or have similar coronary diseases, with the aim of identifying risk factors related to this condition that are specific to women, and consequently, obtaining a classification tool. In our results we obtained that the risk factors that generate the greatest impact on women according to the various sources used in the machine learning models are: last degree of studies, if you are a current smoker, how many cigarettes per day you use, if you suffered from a cerebrovascular event, if you suffer from hypertension (including medicated hypertension), your body mass index, your age, the type of chest pain, blood glucose levels, if you had gestational diabetes, if you have kidney disease, teeth removed, hours of sleep, ethnic group, height in meters, suffering from HIV, suffering from skin cancer, suffer Of alcoholism, having been vaccinated for pneumonia and the flu. Our different models obtained an accuracy of 86%, 100%, 70%, 86% and 100%. We emphasize the importance of incorporating a greater number of risk factors in women for future studies, this will increase the accuracy of specific models in the classification of women at risk of cardiovascular disease and will allow an even better understanding of the impacts of each of the specific risk factors in women.

## I. Introducción

Las cardiopatías son la principal causa de muertes en el mundo; durante los últimos 20 años, la mortalidad incrementó de 2 millones en el 2000, hasta cerca de 9 millones en el 2019, representando el 16% de todas las muertes a nivel mundial (WHO, 2020b) y ocupando las dos principales causas de defunción: la cardiopatía isquémica y el accidente cerebrovascular. (WHO, 2020a) ambas afecciones relacionadas con enfermedades de trastornos trombóticos, vasculitis o enfermedades circulatorias.

A lo largo del tiempo, la enfermedad de las arterias coronarias (EAC) ha sido objeto de estudio, debido al reconocido impacto que una lesión en una arteria coronaria puede tener en la expectativa de vida de una persona. Las observaciones hechas por el Médico James Bryan Herrick fueron pioneras para el estudio de esta patología. Sus descubrimientos propusieron que la obstrucción de una arteria coronaria tenía otras afecciones distintas a una muerte inmediata, que según él mismo relata, era un evento reconocido como “uniformemente fatal”. Sus estudios estuvieron basados en la meticulosa investigación de autopsias, documentando la trombosis como el mecanismo del infarto de miocardio (IM) (Herrick, 1912).

En años posteriores, el médico James B. Herrick continuó documentando como las EAC eran mal diagnosticadas y, como lo menciona en su artículo, “confundidas con problemas gastrointestinales” (Herrick, 1919, p. 388), además, sugiere el uso del electrocardiograma (EKG) para el reconocimiento del IM (Herrick, 1919).

Las aportaciones del médico James B. Herrick fueron pioneras en el descubrimiento de la etiología del IM, siendo en los años previos a sus publicaciones una enfermedad vastamente desconocida. Sus observaciones ayudaron a establecer la formación de coágulos en las arterias coronarias como el mecanismo pivote para el IM, allanando camino para la terapia anticoagulante como tratamiento contra el infarto, sin embargo sus estudios fueron realizados exclusivamente en hombres y animales, descartando el caso clínico específico para la mujer (Wenger, 2012, p. 2).

Las enfermedades cardiovasculares (ECV) son la principal causa de muerte en mujeres en Estados Unidos y en el mundo (Mehta et al., 2016, p. 2), principalmente debido a

TESIS TESIS TESIS TESIS TESIS

las diferencias en los factores de riesgo; factores ginecoobstétricos, la ausencia de actividad física, consumo de tabaco, factores fisiológicos y una menor producción de estrógenos en etapa menopáusica ("5. Sex Affects Health," 2001, p. 164).

En edades avanzadas, específicamente en la edad menopáusica, existe un incremento en la cantidad de la lipoproteína de baja densidad (LDL) (Colesterol malo) y disminución de la lipoproteína de alta densidad HDL (colesterol bueno) en sangre (Revueltas-Jiménez et al., 2022, p. 274), lo que aumenta el riesgo de la aparición de aterosclerosis, condición en la que una sustancia llamada placa, se acumula en las arterias, ésta ocasiona una disminución en el flujo de sangre al corazón (NIH, 2022b). Como consecuencia, una mayor disposición a tener la EAC y una mayor aparición de coágulos (Nielsen, 1995), esto debido a la ausencia de mecanismos protectores en edades avanzadas donde en la etapa menopáusica inhabilita el factor protector de los estrógenos (Lip et al., 1997).

El IM en mujeres, a pesar de su menor incidencia en comparación con los hombres en edades tempranas, puede acontecer incluso durante el período de protección que proporcionan los estrógenos. Esta condición, históricamente menos investigada, ha dado lugar a diagnósticos erróneos y a un tratamiento inadecuado (Wenger, 2012).

La falta de atención hacia el IM en mujeres ha contribuido a un entendimiento insuficiente de sus síntomas y factores de riesgo, lo que ha llevado a complicaciones que pueden llevar a resultados fatales (Arnstein et al., 1996; Bhati & Wunder, 2022), siendo más propensa la mujer a fallecer en el hospital en comparación con el hombre, esto, de acuerdo a los resultados publicados en la revista española de cardiología en el año 2021 (Sambola et al., 2021). Estos resultados resaltan la urgente necesidad de ahondar en investigaciones con mayor atención clínica especializada para mejorar la detección temprana y la gestión eficaz del IM en mujeres.

Nuestro enfoque para solucionar este problema, es usar una herramienta de apoyo diagnóstico y predictivo utilizando herramientas de aprendizaje automático que permita distinguir síntomas de IM en mujeres, lo que brindará una mejor comprensión de esta enfermedad y una atención temprana.

## II. Planteamiento del Problema

La cardiopatía isquémica es la principal causa de muerte en mujeres en una gran parte de los países industrializados, según un estudio llevado a cabo por profesionales de la Asociación Americana del Corazón en 1997, una de cada nueve mujeres presentará alguna forma de ECV y después de los 65 años, 1 de cada 3 (Mosca et al., 1997). Después del año 2000 se ha observado una disminución en las tasas de mortalidad por ECV en el mundo (Zhang et al., 2021, p. 1), a pesar de esto, las ECV siguen siendo la principal causa de muerte a nivel mundial (WHO, 2020b).

La tasa de mortalidad en mujeres con enfermedad cardíaca reumática y cardiopatía hipertensiva es mayor a la de los hombres, a pesar de que las estadísticas globales muestran una mayor cantidad de hombres que padecen de esta enfermedad (Zhang et al., 2021, pp. 2 - 9).

Un estudio realizado en México encontró que debido a que inicialmente los estudios sobre ECV excluían a la mujer; se ha generado la falsa creencia que la cardiopatía isquémica ocurre casi exclusivamente en los hombres, siendo los síntomas y signos clínicos como la presencia de dolor torácico menos específicos en la mujer, por consiguiente de lo anterior, la mujer presenta una mayor mortalidad por infarto agudo de miocardio (IAM), una de las principales causas de lo anterior es debido a una mayor frecuencia de infarto asintomático, re-incidencia, insuficiencia cardíaca, entre otras enfermedades asociadas (Solorio et al., 2007).

Los factores de riesgo cardiovascular como lo son: el fumar, padecer diabetes, síndrome de antifosfolípidos, entre otros; muestran una tendencia que desfavorece a la mujer, siendo un 25% más alta la mortalidad en mujeres fumadoras que entre hombres fumadores (Wenger, 2012, pp. 10 - 13).

La participación de mujeres en estudios relacionados con ECV es muy baja. En Europa, su participación en estudios clínicos rondaba entre el 16% y el 25%, mientras que para Estados Unidos, el 46% de pacientes con ECV son mujeres y tan sólo el 25% participó en estudios para la prevención de ECV (Wenger, 2012, p. 14).

TESIS TESIS TESIS TESIS TESIS

También se reporta que en la mujer prevalece más la angina de pecho, síntoma de una mala irrigación al miocardio, teniendo períodos más largos de isquemia en EKG, cuyos efectos resultaban con consecuencias graves (Wenger, 2012, p. 16). En el mismo artículo publicado por la Asociación Americana del Corazón (AHA) se presenta que mujeres con un IM suelen ser recurrentes, exhortando que se continúe con la investigación de esta patología (Wenger, 2012, p. 17).

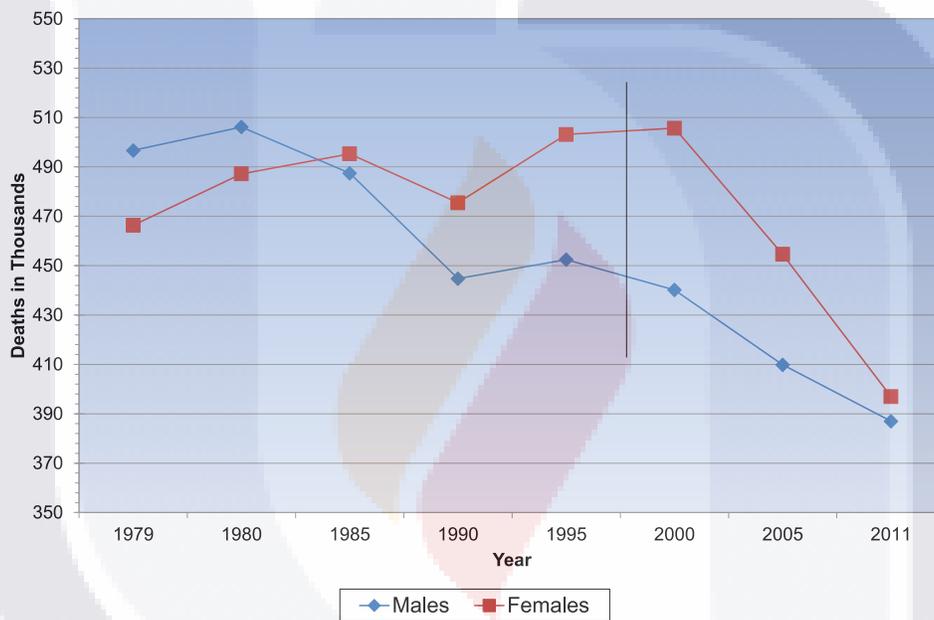
La carga excesiva de trabajo, problemas financieros, depresión y padecer de estrés crónico, según un estudio del estado del arte (Cho et al., 2020), son algunos factores de riesgo reconocidos por la comunidad científica que afectan especialmente a la mujer. Además, factores emergentes en la mujer, que se dan dependiendo de su edad, hábitos alimenticios, factores hormonales, entre otros (Mulvagh et al., 2022) subrayan la necesidad de abordar la salud de la mujer desde una perspectiva integral que involucre tanto factores de riesgo tradicionales, como factores de riesgo exclusivos en la mujer.

Las herramientas computacionales ofrecen numerosas ventajas para la resolución de problemas. Entre las más destacadas se encuentran el apoyo en la toma de decisiones y la capacidad de proporcionar soluciones rápidas y eficientes. Sin embargo, existe un desafío significativo relacionado con la baja participación de mujeres en estudios epidemiológicos y clínicos vinculados con la salud cardiovascular. Esta falta de representación dificulta el acceso a conjuntos de datos que reflejen de manera precisa los factores de riesgo específicos de la mujer.

Para abordar este problema, es fundamental promover una mayor inclusión de mujeres en estos estudios, lo que no sólo mejoraría la calidad de los datos disponibles, sino que también permitiría el desarrollo de soluciones más equitativas y personalizadas en el ámbito de la salud. Sin una representación adecuada, las herramientas computacionales pueden caer en sesgos y ofrecer soluciones que no consideren las necesidades particulares de las mujeres.

### III. Justificación

El continente americano no es una región homogénea; la incidencia de mortalidad no sigue una tendencia concreta (Arroyo-Quiroz et al., 2020), Las mujeres con síndrome isquémico coronario agudo (SICA) que viven en países con altos estándares socioeconómicos tienen un menor riesgo de mortalidad después del alta hospitalaria en comparación con los hombres, Sin embargo, en países con un contexto socioeconómico más bajo, la mortalidad ajustada entre mujeres y hombres es similar (Rossello et al., 2021).

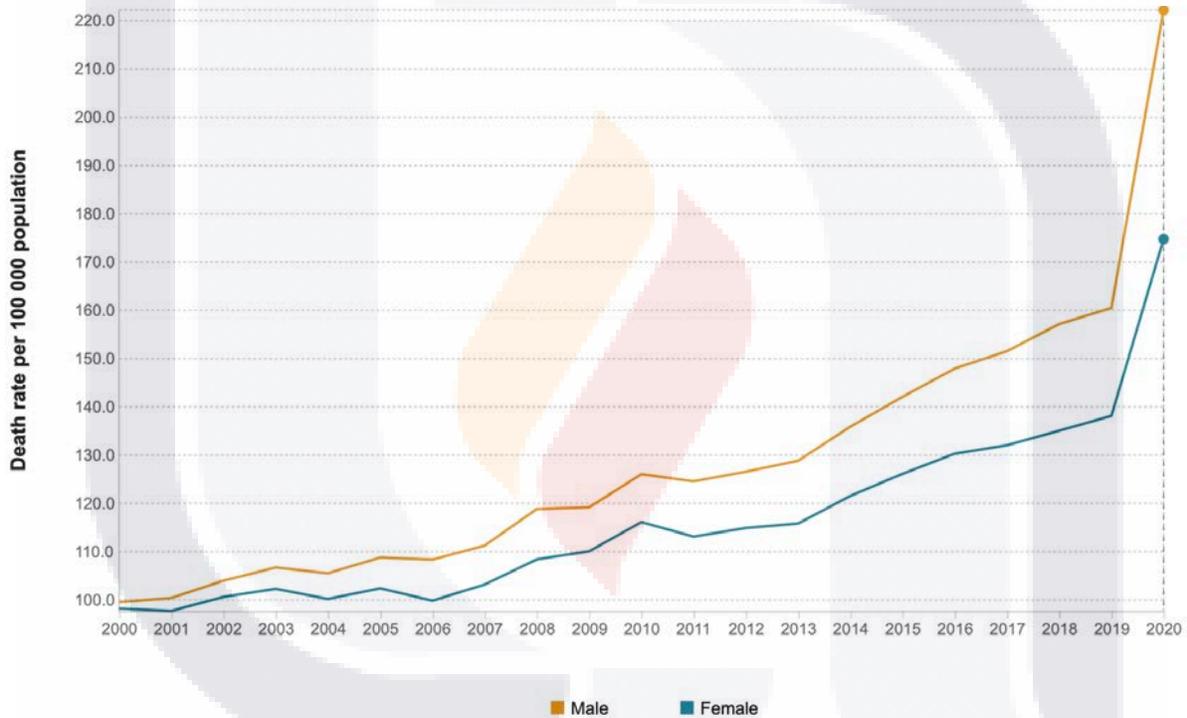


**FIG 1.** Mortalidad por ECV para hombres y mujeres en Estados Unidos de 1979 a 2011 (Mozaffarian et al., 2015).

En la figura 1 se observa una disminución promedio anual del 4.4% entre 2001 y 2011 (Mozaffarian et al., 2015). Esta reducción se atribuye a las distintas campañas de concientización, al uso de terapias basadas en la evidencia y a las intervenciones en mujeres con factores de riesgo de ECV (Mehta et al., 2016, p. 9). Sin embargo, desde 1984, cada año mueren más mujeres por enfermedades cardiovasculares que hombres. En particular, las muertes por EAC entre mujeres superan a las causadas por todos los tipos de cáncer combinados (Wenger, 2012, p. 3).

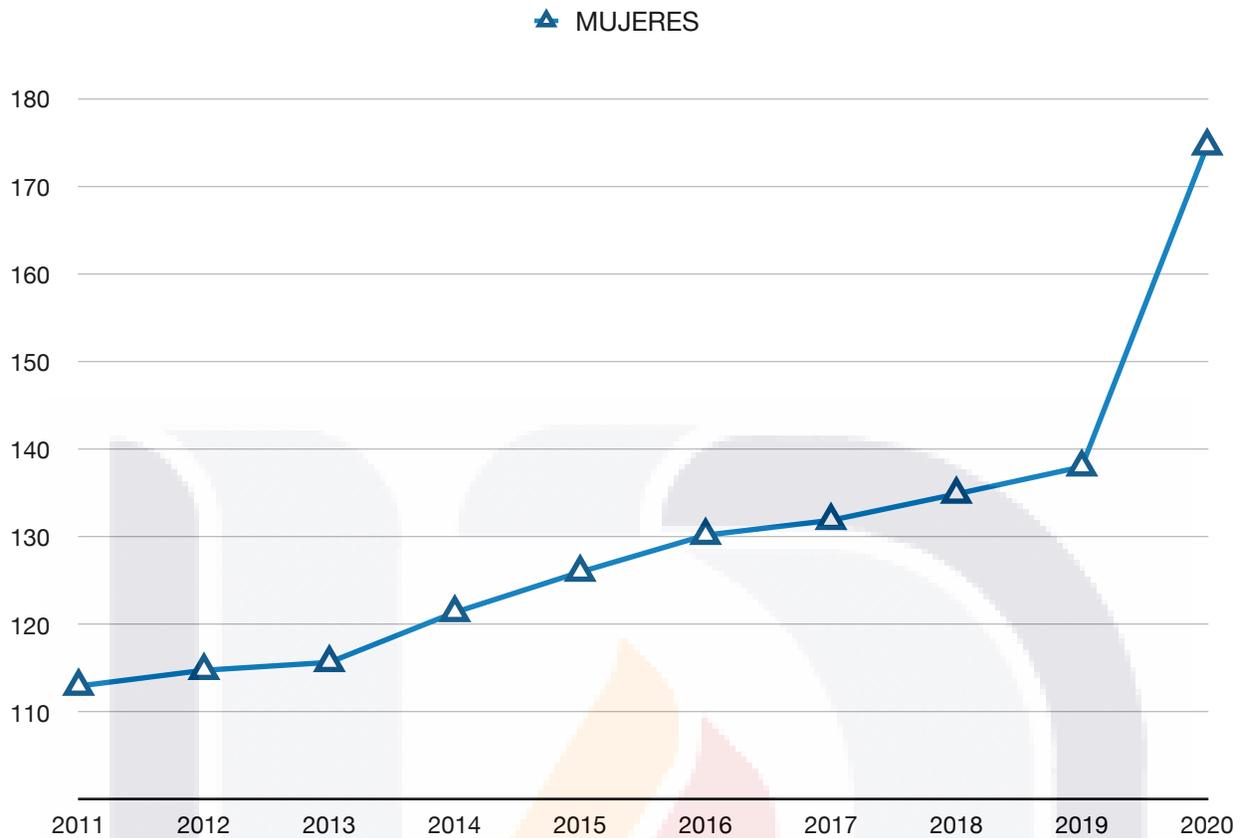
### Efectos en la comunidad

A diferencia de Estados Unidos y Canadá, Latinoamérica no presenta una incidencia homogénea de EAC (Pagan et al., 2017, p. 1). Entre el 2001 y 2013 la tasa de mortalidad por EAC en mujeres se redujo en Ecuador un 59%, en Uruguay un 51% y en Colombia y Cuba un 35%. En contraste, la tasa de mortalidad por EAC en mujeres aumentó en Panama y México, con incrementos del 52% y del 4% respectivamente (Arroyo-Quiroz et al., 2020, p. 1). El incremento en la mortalidad por ECV también se refleja en la base de datos de la Organización Mundial de la Salud, que muestra un aumento de la ECV en este país (WHO, 2022) (FIG. 2).



**FIG 2.** Tasa de mortalidad de ECV por cada 100,000 habitantes por sexo en Mexico. (WHO, 2022).

El Sector Salud en México menciona que una de cada cinco personas pueden fallecer debido a una ECV, lo cual convierte a esta enfermedad en un gran desafío. Las ECV representan el 20.2% de los fallecimientos relacionados con problemas de salud (Salud, 2018), mostrando índices en ascenso (FIG. 2).



**FIG 3.** Muertes para las mujeres y grupo de edad en Mexico por ECV. (WHO, 2022).

Entre el 2000 y el 2020, la tendencia de mortalidad por ECV en mujeres mexicanas incrementó un 83% (FIG. 3). Debido a la pandemia de COVID-19, durante el periodo de 2019 a 2020, la tendencia incremento, asociándose con un 51.2% de los casos de mortalidad (Bader et al., 2021).

En México, en 1990, las enfermedades isquémicas del corazón ocupaban el noveno puesto como causa de muerte para ambos sexos y el sexto puesto para las mujeres. Para el año 2010, las enfermedades isquémicas del corazón se convirtieron en la tercer causa de muerte prematura en mujeres (Lozano et al., 2013, p. 586).

Entre el 2011 y el 2020 la tendencia de mortalidad por ECV en mujeres ha incrementado un 4% (WHO, 2022) (FIG. 2), mientras que para el año 2020, la principal causa de muerte en mujeres fue la ECV, superando las 97,000 defunciones, mientras que para ese mismo año hubieron 70,000 muertes causadas por COVID-19, lo cual

indica un alto índice de fallecimientos relacionados con las ECV en mujeres mayores a los 45 años, siendo la Ciudad de México la ciudad con mayor tasa de defunciones (INEGI, 2020).

En el estado de Aguascalientes, este problema de salud ocupa el segundo lugar en la tabla de mortalidad según cifras del año 2021, en el que el número de defunciones alcanzó los 9,712 fallecimientos, siendo el coronavirus la principal causa de muerte en dicho estado mexicano con 2,089 muertes y en segundo lugar las enfermedades del corazón, con 1,538 muertes, como puede observarse en la Tabla 1 (INEGI, 2021).

TABLA 1. México: principales causas de mortalidad en el estado de Aguascalientes 2021. Fuente: (INEGI. Recuperado el 31 de Marzo del 2023, desde inegi.org.mx)

<b>Causa del fallecimiento</b>	<b>Numero de fallecimientos</b>
<b>COVID-19</b>	2,089
<b>Enfermedades del corazón</b>	1,538
<b>Tumores malignos</b>	945
<b>Diabetes mellitus</b>	867
<b>Influenza y neumonía</b>	494
<b>Accidentes</b>	417
<b>Enfermedades cerebrovasculares</b>	337
<b>Enfermedades del hígado</b>	275
<b>Enfermedades pulmonares obstructivas crónicas</b>	254
<b>Lesiones autoinfligidas intencionalmente</b>	174
<b>Desnutrición y otras deficiencias nutricionales</b>	134
<b>Ciertas afecciones originadas en el periodo perinatal**</b>	123
<b>Insuficiencia renal</b>	113
<b>Agresiones</b>	99
<b>Malformaciones congénitas, deformidades y anomalías cromosómicas</b>	88
<b>Síndrome de dependencia del alcohol</b>	69

Las ECV son una problemática a nivel mundial (WHO, 2022) y en México (Salud, 2018), específicamente en la mujer, debido a que sus síntomas llegan a ser atípicos y menos reconocibles en comparación con los hombres, incrementando el riesgo de no recibir un tratamiento adecuado. Según el estudio “Impact of initial hospital diagnosis on mortality for acute myocardial infarction” los autores reconocen que “las mujeres con infarto de miocardio con elevación del segmento ST tenían un 59% más de probabilidades de un diagnóstico erróneo inicial en comparación con los hombres” (Wu et al., 2018).



## IV. Objetivos

### Objetivo General

El objetivo de este estudio es desarrollar un clasificador usando aprendizaje automático, mediante el cual se pueda detectar de manera temprana la tendencia a desarrollar una ECV en mujeres, atendiendo los posibles síntomas definidos y no definidos.

### Objetivos Específicos

- Implementar varios clasificadores mediante aprendizaje automático para determinar cuál se adapta mejor a la problemática de ECV en mujeres.
- Identificar factores de riesgo vinculados con cardiopatías isquémicas en mujeres utilizando una herramienta de aprendizaje automático.
- Desarrollar una aplicación la cual ayude al cuerpo médico a diagnosticar mujeres en riesgo de padecer una ECV.

Se pretende desarrollar una herramienta para ayudar al personal de salud a identificar a mujeres en riesgo de padecer ECV, permitiendo que el equipo médico pueda dirigir una mayor atención y recursos hacia aquellas pacientes con alto riesgo de infarto.

### Preguntas de Investigación

¿Es posible clasificar el infarto en las mujeres utilizando una herramienta computacional de aprendizaje automático?

- ¿Que modelo de aprendizaje automático se adecúa mejor para desarrollar una herramienta de clasificación del infarto en la mujer?
- ¿Es posible identificar factores de riesgo de infarto en mujeres haciendo uso de una herramienta computacional?
- ¿Es posible clasificar el infarto en las mujeres únicamente con los factores de riesgo tradicionales?

## V. Marco Contextual

Distintas investigaciones relacionadas con la valoración y tratamiento de cardiopatías isquémicas se han llevado a cabo haciendo uso de herramientas computacionales y equipos tecnológicos: las herramientas que permiten la predicción del impacto de un infarto son desarrolladas y utilizadas principalmente para órganos vitales como el cerebro y el corazón. Debido a la importancia que tiene mantener el tejido del cerebro oxigenado durante una intervención quirúrgica, la mayoría de estas investigaciones están realizadas alrededor de los procedimientos relacionados con el cerebro.

Hay un creciente interés en el uso de algoritmos para la predicción de cardiopatías, en el estudio “Infarct prediction and treatment assessment with MRI-based algorithms in experimental stroke models” proponen el uso de múltiples modalidades de resonancia magnética (MRI) para predecir el infarto de tejido en un accidente cerebrovascular humano agudo. En su estudio desarrollaron y probaron un algoritmo de modelo lineal generalizado (GLM) basado en voxels para predecir el infarto en el tejido de un animal en un modelo de accidente cerebrovascular, con el fin de comparar directamente el resultado predicho con el resultado histológico del tejido y para evaluar el potencial de evaluar la eficacia terapéutica utilizando estos algoritmos multiparamétricos. El resultado previo al tratamiento en comparación con la histología posterior al tratamiento fue muy preciso en ratas tratadas con solución salina. Además, “las alteraciones en la progresión de la enfermedad debido al tratamiento se controlaron con mayor sensibilidad con sus técnicas de análisis basadas en voxel que con enfoques volumétricos. El estudio muestra que los algoritmos predictivos son métricas prometedoras para el diagnóstico, el pronóstico y la evaluación terapéutica después de un accidente cerebrovascular agudo que pueden traducirse fácilmente de entornos preclínicos a clínicos” (Wu et al., 2007). Este primer estudio se realizó utilizando herramientas costosas, por lo que llevar a cabo una investigación de este tipo es poco viable. Por otro lado, se tiene el análisis holístico para la predicción de un infarto al miocardio: En el artículo “Efficient heart disease prediction system using decision tree”, el autor propuso un sistema que puede predecir automáticamente las enfermedades cardíacas. El objetivo principal de esta investigación fue la transformación de datos en

TESIS TESIS TESIS TESIS TESIS

información útil que pueda permitir a los profesionales de la salud tomar decisiones clínicas efectivas. Utilizaron el conjunto de datos de enfermedades cardíacas; diseñaron un sistema que utiliza un árbol de decisión como un método que puede descubrir de manera eficiente las reglas para predecir el nivel de riesgo de los pacientes en función del parámetro dado sobre su salud. Las reglas se pueden priorizar en función de los requisitos de los usuarios. Habían utilizado una herramienta para la Extracción del conocimiento basada en el aprendizaje evolutivo (KEEL) (Purushottam et al., 2015). Además, en el artículo "A Self-Test to Detect a Heart Attack Using a Mobile Phone and Wearable Sensors", el investigador propuso una aplicación de teléfono móvil que puede ayudar a los pacientes a identificar si están teniendo un ataque al corazón o no sin acudir a un especialista en persona. Han utilizado una aplicación de teléfono móvil con algunas preguntas para analizar, un sensor de electrocardiograma (ECG) desgastado, un dispositivo de medición de la presión arterial. Si una persona está en peligro y no puede llamar a una ambulancia, el teléfono móvil determinará automáticamente la ubicación actual de la persona que utiliza WiFi, GSM o GPS y enviará mensajes de voz y de texto automatizados a su cardiólogo (Leijdekkers & Gay, 2008).

En el artículo "Analysis of coronary heart disease and prediction of heart attack in coal mining regions using data mining techniques", el autor analizó la tasa de ECV en las regiones mineras de carbón de Singareni en el estado de Andhra Pradesh e India. Han utilizado técnicas de minería de datos: Decision Trees, Naive Bayes y Neural Network como método y un conjunto de datos del repositorio UCI para su análisis. El modelo bayesiano (BN) logró una precisión de clasificación de 0,82 con una sensibilidad de 0,87, los árboles de decisión (C4.5) lograron una precisión de clasificación de 0,825 con una sensibilidad de 0,8717. Sin embargo, el modelo de red neuronal (MLP) mostró el mejor de los cuatro modelos evaluados. MLP logró una precisión de clasificación de 0,897 con una sensibilidad de 0,9017 (Srinivas et al., 2010). Los artículos revisados que realizan una predicción del infarto utilizando cuestionarios no toman en cuenta los factores específicos en la mujer como puede ser la depresión, factores hormonales, consumo de medicamentos, niveles de estrés, factores psicosociales, etc. que en el caso de la mujer podrían brindar información relevante para la enfermedad.

## VI. Marco Teórico

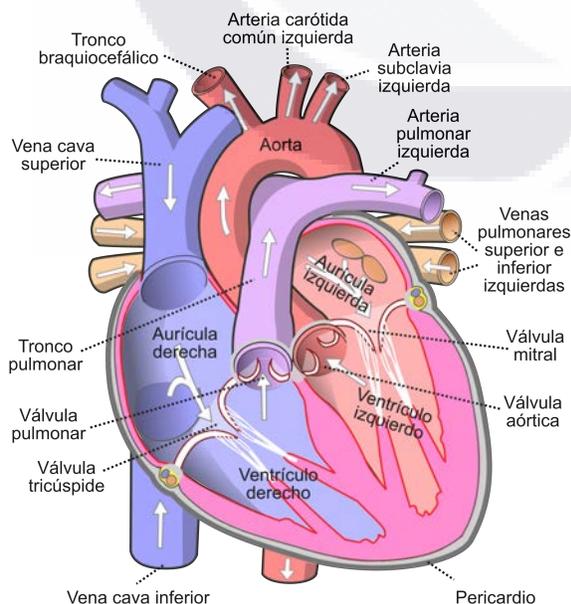
El corazón es un órgano vital para el cuerpo humano, por medio de la sangre suministra oxígeno y nutrientes a los tejidos, transporta el dióxido de carbono y se encarga de transportar productos de deshecho a diferentes órganos como los pulmones, que hacen el intercambio de gases (hematosis) (NIH, 2022a).

### Sistema cardiovascular

Los componentes del sistema cardiovascular son: el corazón, los vasos sanguíneos y la sangre, sus principales funciones son: transportar oxígeno y nutrientes a los tejidos, transportar bióxido de carbono y productos de deshecho metabólicos de los tejidos a los pulmones y a los órganos excretores, distribución de agua, electrolitos y hormonas, medio de contribución al sistema inmune y la termorregulación (Aaronson et al., 2020).

### Anatomía del corazón

El corazón se divide en 4 cámaras, dos del lado derecho y dos del lado izquierdo, en el lado superior del corazón encontramos dos aurículas o atrios, y dos inferiores, denominadas ventrículos, estos ventrículos se encuentran separados mediante un sistema de válvulas que evitan el retorno de flujo sanguíneo, existen dos tipos de válvulas, las atrio-ventriculares (mitral y tricúspide) y las semilunares (pulmonar y aórtica) (Saxton et al., 2023). Esto puede ser visto en la Figura 4.



**FIG 4.** Diagrama del corazón humano. (Rhcastilhos, 2010).

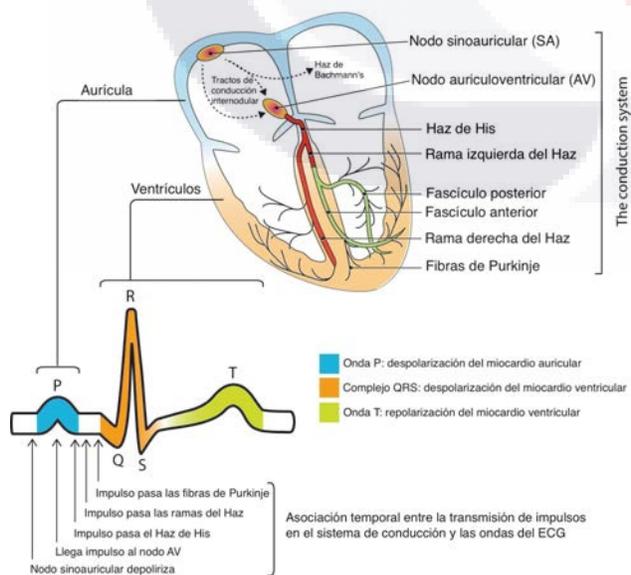
El corazón está principalmente compuesto de miocitos, estas células de tipo muscular tienen la función de contraerse y de transmitir el impulso eléctrico producido por el sistema eléctrico del corazón (Saxton et al., 2023).

### Sistema eléctrico del corazón

Cada latido del corazón es producido por un conjunto de células especializadas, los nodos sinoauricular y atrioventricular. La contracción comienza en las aurículas mediante el impulso producido en el nodo sinoauricular, enseguida el impulso eléctrico llega al nodo atrioventricular, pasa por el Haz de His a las diferentes ramas hasta llegar a las fibras de Purkinje, las cuales comienzan la contracción de los ventrículos (sístole atrial y ventricular). Por ultimo se da la relajación de las cámaras atriales y ventriculares (diástole) (Figura 5) (Richard N. Fogoros, 2023).

El descubrimiento de este sistema eléctrico se dió en 1856 en Alemania por los científicos Kölliker y Müller por lo que en 1869 se realizó el primer electrocardiograma en un ser humano por Muirhead en Londres. Posteriormente, este descubrimiento permitió realizar el diagnóstico de Isquemias de miocardio y EAC (Johansson, 2001).

El uso del EKG (Figura 4) mide la actividad eléctrica producida en las pulsaciones del corazón, en el EKG mostrará la velocidad y el ritmo de las contracciones en las cámaras superiores e inferiores del corazón (AHA, 2022).



**FIG 5.** Diagrama del sistema eléctrico del corazón. (Learning, 2023).

## **Enfermedades del corazón**

Existen distintas ECV, entre ellas tenemos: cardiopatía congénita, trastornos del ritmo cardiovascular, aterosclerosis sub-clínica, enfermedad coronaria, insuficiencia cardíaca, enfermedad valvular y la enfermedad arterial periférica y el infarto agudo de miocardio (IAM) (Tsao et al., 2022).

Una de las ECV más relacionadas con el IAM es la Angina Inestable, la cual es caracterizada por el dolor de pecho, garganta y acompañada por dificultad para respirar, náuseas, mareos, sudoración, ritmo cardíaco notable, entre otros y se produce debido a una restricción en la irrigación sanguínea al corazón.

## **Mecanismo del IAM**

Las EAC son causadas por el estrechamiento de las arterias del corazón (coronarias), las cuales proveen oxígeno al músculo del corazón (miocardio) ("Ischemic Heart Disease," 2010). Este estrechamiento o disminución de flujo de sangre en las arterias coronarias comúnmente es causado por la acumulación de placa, evento llamado aterosclerosis, produciendo un daño en las arterias que impide el flujo correcto de sangre (trombosis) (Ashorobi et al., 2022), aunque también puede ser causado por un coágulo de sangre, un émbolo o por una constricción de los vasos sanguíneos (Libby & Theroux, 2005). Lo cual ocasiona la muerte del tejido del miocardio (necrosis), por lo cual deja de funcionar, produciendo el IAM (Prakash C Deedwania, 2022).

El IAM es una forma grave de ECV que a menudo es difícil de diagnosticar sin el uso de equipo médico especializado. Sin embargo, conocer el riesgo de padecer una ECV y disponer una herramienta para medir el riesgo de sufrir IAM pueden ser de gran ayuda para el equipo médico, aportando un alto significativo en la detección temprana de de estas afecciones.

## Factores de riesgo de enfermedad cardiovascular

Entre los factores de riesgo no modificables para la enfermedad coronaria se incluyen la edad, el género, el origen étnico y el historial familiar de EAC, los factores de riesgo modificables incluyen hipertensión, hiperlipidemia, diabetes, obesidad, fumar, tener una mala dieta, un estilo de vida sedentario, y el estrés (Brown et al., 2020).

La Universidad de California San Francisco (UCSF) reconoce los siguientes factores de riesgo no modificables para las enfermedades cardíacas: edad, género, antecedentes familiares de enfermedades cardíacas y la raza. Y los factores de riesgo modificables los siguientes: hipertensión, fumar, colesterol, colesterol de baja densidad, colesterol de alta densidad, colesterol total, triglicéridos, diabetes y pre-diabetes (Grossman, 2024).

En esta publicación universitaria hace una breve descripción de los factores de riesgo mostrados en la Tabla 2.

TABLA 2. Factores de riesgo modificables y no modificables. (Grossman, 2024).

Factor de Riesgo	Descripción	Tipo
<b>Edad</b>	Según los cálculos de la Asociación Americana del Corazón, alrededor del 80 por ciento de las personas que mueren por ECV tienen 65 años o más.	No Modificables
<b>Genero</b>	Durante mucho tiempo se ha considerado que las enfermedades cardíacas son principalmente enfermedades de los hombres, pero el resultado en mujeres a menudo es peor.	
<b>Antecedentes Familiares de ECV</b>	Su riesgo de desarrollar una enfermedad cardíaca aumenta si tiene un familiar que desarrolló una enfermedad cardíaca a una edad temprana (antes de los 55 años).	
<b>Raza</b>	Los afroamericanos corren un mayor riesgo de desarrollar ECV.	
<b>Hipertensión</b>	Muchas personas tienen presión arterial alta durante años. Si no se trata, puede provocar una enfermedad coronaria, un ataque cardíaco o un accidente cerebrovascular.	Modificables
<b>Fumar</b>	Los fumadores tienen más del doble de riesgo de desarrollar ECV. En promedio, fumar reduce la expectativa de vida 13 años para un fumador masculino y 14 años para una mujer fumadora.	
<b>Colesterol de baja densidad</b>	La lipoproteína de baja densidad contribuye a los bloqueos de las arterias (placas).	

Factor de Riesgo	Descripción	Tipo
<b>Colesterol de alta densidad</b>	Los altos niveles de colesterol HDL reducen el riesgo de desarrollar ECV.	Modificables
<b>Triglicéridos</b>	Es el tipo de grasa más común en el cuerpo. Muchas personas que tienen enfermedades cardíacas o diabetes tienen niveles altos de triglicéridos. El nivel normal de triglicéridos es inferior a 150 mg/dL.	
<b>Colesterol total</b>	Es una medida del colesterol LDL, el colesterol HDL y otros lípidos. El nivel deseable de colesterol total es inferior a 200 mg/dL.	
<b>Diabetes</b>	La diabetes (nivel elevado de azúcar en la sangre) aumenta el riesgo de desarrollar ECV.	
<b>Pre-diabetes</b>	Si no hacen cambios en el estilo de vida, es probable que desarrollen diabetes en los próximos 10 años.	

De estos factores de riesgo, el autor comenta información importante acerca de algunos de ellos:

**El genero:** Aunque las mujeres tienden a desarrollar ECV unos 10 años más tarde en la vida que los hombres, el resultado para las mujeres es a menudo peor: 2/3 de las mujeres estadounidenses tienen al menos 1 factor de riesgo, Más del 60 % tiene sobrepeso u obesidad, El 25 % casi NO tiene actividad física, Más del 50 % de las mujeres mayores de 45 años tienen hipertensión, El 40 % de las mujeres mayores de 55 años tienen colesterol elevado, El HDL bajo (menos de 40 mg) es un factor de riesgo más fuerte para las mujeres mayores que para los hombres mayores, La diabetes aumenta el riesgo relativo de 3 a 7 veces para las mujeres (Grossman, 2024).

**Antecedentes familiares de ECV:** Si tus padres desarrollaron una ECV más adelante en la vida, puede estar relacionada con la edad en lugar de ser genética. Si bien no puede cambiar sus genes, es importante conocer su historial médico familiar y compartirlo con su médico (Grossman, 2024).

**Hipertensión:** Un porcentaje más alto de hombres que mujeres tiene hipertensión hasta los 45 años. De 45 a 64 años, el porcentaje de hombres y mujeres es similar. Después de eso, un porcentaje mucho mayor de mujeres tienen hipertensión que los hombres (Grossman, 2024).

TESIS TESIS TESIS TESIS TESIS

Fumar es el factor de riesgo más prevenible. La exposición al humo (fumadores pasivos) aumenta el riesgo incluso para los no fumadores. alrededor del 20% de adultos mayores de 18 años en los EE. UU. son fumadores (Grossman, 2024).

Colesterol de baja densidad: La mayoría de las personas deben apuntar a un nivel de colesterol LDL de 100 mg/dL o inferior. Si tiene un riesgo muy alto de desarrollar ECV, o si ya ha tenido un ataque al corazón, es posible que deba apuntar a un nivel de LDL por debajo de 70 md/dL (Grossman, 2024).

Colesterol de alta densidad: La lipoproteína de alta densidad es una proteína de transporte inverso: elimina el colesterol de las arterias y lo lleva al hígado, donde se puede transmitir fuera del cuerpo (Grossman, 2024).

Diabetes: La diabetes mellitus se define como una glucosa en sangre en ayunas superior a 125 mg/dL. Hay dos tipos principales de diabetes. La diabetes tipo 1 generalmente se diagnostica por primera vez en niños y adultos jóvenes. La diabetes tipo 2 es la forma más común y puede desarrollarse a cualquier edad. Si tienes diabetes, sin importar el tipo, significa que tienes demasiada glucosa en la sangre y puede provocar graves problemas de salud. La diabetes y las enfermedades cardíacas comparten factores de riesgo similares (nivel alto de colesterol, presión arterial alta, obesidad) (Grossman, 2024).

Pre-Diabetes: Las personas con un nivel de glucosa en sangre en ayunas entre 100 mg/dL y 125 mg/dL tienen un mayor riesgo de desarrollar diabetes tipo 2. La pre-diabetes es reversible. Si la persona afectada pierde peso, mantiene una dieta saludable y aumenta su actividad física, puede ser capaz de prevenir la progresión a la diabetes (Grossman, 2024).

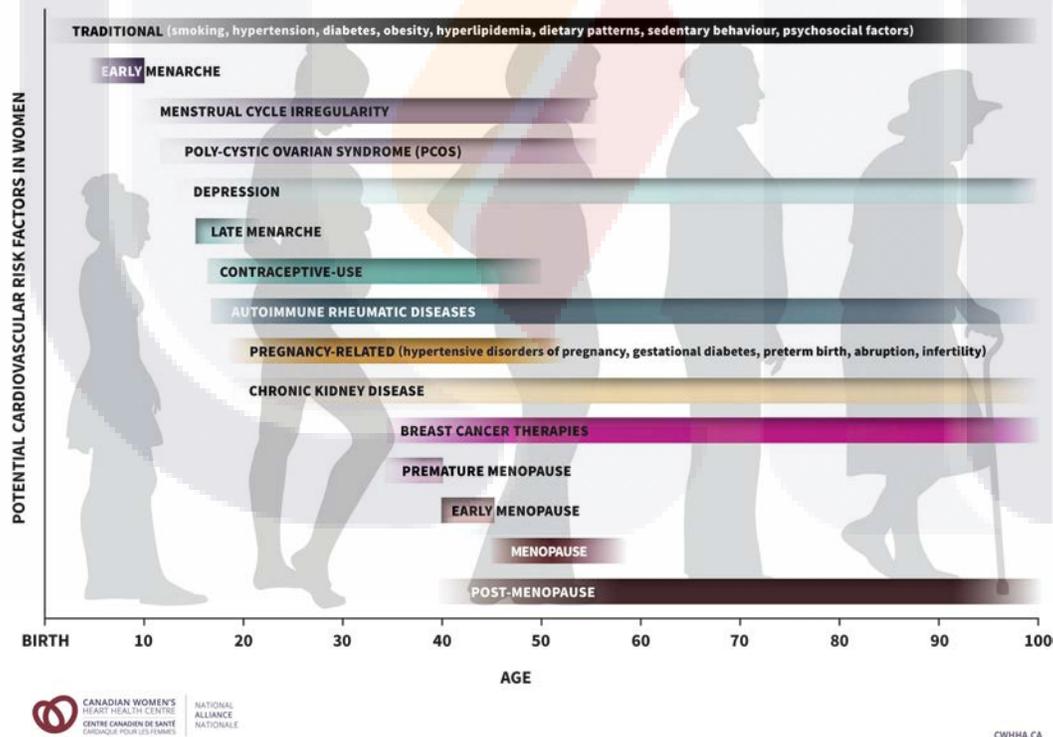
Se les puede llamar cardiopatías o enfermedades cardíacas a las enfermedades del corazón, Las ECV tienen que ver con las arterias coronarias o los vasos sanguíneos del corazón (Grossman, 2024).

Distintos estudios logran identificar diferentes factores de riesgo modificables y no modificables, haciendo un resumen: encontramos los siguientes factores de riesgo tradicionales y emergentes en la mujer (Grossman, 2024).

## Factores de riesgo en mujeres

En la literatura para el caso específico de la mujer se mencionan distintos factores de riesgo, rasgos fisiológicos y factores ginecoobstétricos, que son determinantes ante la adquisición de la enfermedad, a estos factores que son específicos para la mujer se les conoce como factores de riesgo emergentes (Vancheri et al., 2022).

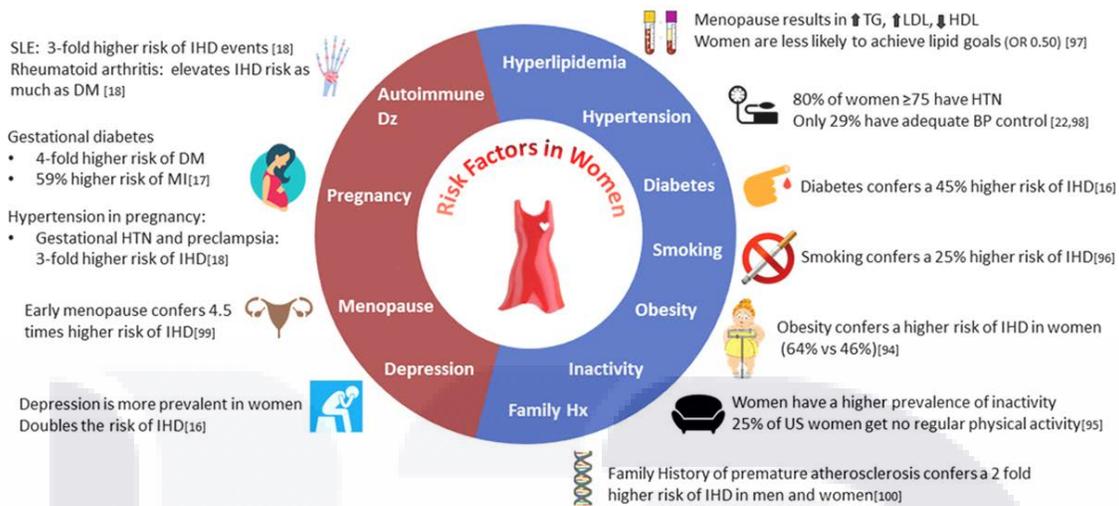
De acuerdo al Centro Canadiense de Salud del Corazón de la Mujer (Canadian Women’s Heart Health Centre) los factores de riesgo tradicionales siempre están presentes durante la vida de la mujer, mientras que los factores de riesgo emergentes cambian en función de la edad de la mujer (FIG. 6), como lo es en el caso de tener una edad menarca temprana, padecer síndrome de ovario poliquístico, depresión, enfermedad reumática autoinmune, enfermedades relacionadas con el embarazo, enfermedad de los riñones, terapia del cáncer de senos, menopausia prematura, depresión, entre otros (Mulvagh et al., 2022).



**FIG 6.** Edad promedio y el periodo de factores de riesgo tradicionales y emergentes (Mulvagh et al., 2022).

## Emerging Risk Factors

## Traditional Risk Factors



**FIG 7.** Factores de riesgo tradicionales y emergentes en mujeres (Aggarwal et al., 2018).

Además el estudio “Sex Differences in Ischemic Heart Disease” reconoce que los factores de riesgo tradicionales para la cardiopatía isquémica en mujeres suponen un riesgo añadido para las mujeres en comparación con los hombres (FIG. 7). Los factores de riesgo que se mencionan son la artritis reumatoide, la diabetes gestacional, la hipertensión durante el embarazo, la depresión, entre otras (Aggarwal et al., 2018).

La comunidad científica valora la relación entre los problemas gastrointestinales y los problemas en el corazón (Álvarez Vega et al., 2021). Así como los patrones de sueño como marcadores clínicos para la ECV (Full et al., 2023). Sin embargo estos factores de riesgo no han sido clasificado determinantes para la ECV en la mujer.

Los factores de riesgo emergentes para mujeres de acuerdo con ambos estudios son: edad menarca temprana, irregularidades en el periodo menstrual, síndrome de ovario poliquístico, depresión, edad menarca tardía, uso de anticonceptivos hormonales, enfermedad reumática autoinmune, diabetes gestacional, enfermedad crónica de los riñones, terapias para cancer de pecho, una menopausia temprana y tardía, y el periodo post-menopáusico (Mulvagh et al., 2022). Además de los factores de riesgo como: depresión, hipertensión y diabetes gestacional y la artritis reumatoide también son factores de riesgo emergentes (Aggarwal et al., 2018).

## **Inteligencia artificial**

Dentro de las ciencias de la computación existe un campo en el que se investiga la búsqueda de soluciones a diferentes problemáticas complejas a partir de la simulación de la inteligencia humana, aportando de herramientas que las hace capaces de aprender y resolver problemas imitando el comportamiento humano del proceso de aprendizaje mediante el reconocimiento de patrones en información etiquetada, la agrupación de objetos basándose en sus características, reconocimiento de imágenes y la comprensión del lenguaje natural, lo que permite solucionar problemas complejos y apoyar a la toma de decisiones (Hunt, 2014).

El aprendizaje automático es el conjunto de técnicas del campo de la inteligencia artificial (IA) que permite a las computadoras aprender de una gran cantidad de datos para hacer regresión o clasificación basándose en la información intrínseca de los mismos datos haciendo uso de técnicas estadísticas y diferentes algoritmos dedicados a la generación de modelos (Zhou, 2021). La generación de modelos de aprendizaje automático puede provenir de diferentes algoritmos, cada uno tiene diferentes implementaciones y por la naturaleza del mismo puede ser útiles en ciertos casos, por lo que para escoger los algoritmos a utilizar deberemos hacer una revisión de literatura y encontrar que modelos se adecuan mejor al contexto médico.

En el estudio “Predicción inteligente de la enfermedad clínica basada en la IA utilizando un clasificador de bosque aleatorio y Naive Bayes” (AI-based smart prediction of clinical disease using random forest classifier and Naive Bayes) los autores utilizaron el algoritmo de bosque aleatorio (Breiman, 2001), comparando los resultados distintos conjuntos de datos de: diabetes, EAC, y cancer se obtuvo que el algoritmo de bosque aleatorio superó a “Naive Bayes”, obteniendo una puntuación para el modelo entrenado con los datos de EAC del 83.85% (Jackins et al., 2021). Además, los autores del artículo “Técnica de regresión logística para la predicción de ECV” (Logistic regression technique for prediction of cardiovascular disease) encontraron que un modelo que hace uso de regresión logística (Seber & Lee, 2003) obtuvo un 87.10% de exactitud para la clasificación de personas que se encontraban en la unidad de cuidados intensivos (G et al., 2022).

La publicación del estudio “Predicción de ECV utilizando máquinas de vectores de apoyo” (Cardiovascular disease prediction using support vector machines) donde el uso de máquinas de soporte vectorial (Suthaharan, 2016) ayudó a determinar si una persona sufre de rigidez arterial “y, por lo tanto, el riesgo de desarrollar ECV” con una exactitud mayor al 85% (Alty et al., 2003). Además, otro estudio utilizó máquinas de soporte vectorial para determinar si un paciente padece ECV, en sus resultados lograron una exactitud del 86%, la cual mencionan es superior a otros sistemas de producción encontrados en la literatura (Rani et al., 2021).

## Bosque aleatorio

Es un algoritmo de aprendizaje supervisado utilizado para la solución de problemas de clasificación y de regresión. Se basa en el concepto de “aprendizaje por conjuntos” (Ensemble Learning) mediante el cual obtiene el resultado de múltiples modelos de aprendizaje, en este caso múltiples árboles de decisión, mejorando la precisión del resultado. Una de las características principales de este algoritmo es la selección aleatoria de características, durante la construcción de cada árbol de decisión, para cada división de nodo, el algoritmo escoge un subconjunto aleatorio de características predictoras, lo cual mejora la generalización y reduce el sobre-ajuste. Primero genera una colección de árboles de decisión independientes, enseguida genera un muestreo

por reemplazo, seleccionando de forma aleatoria muestras del conjunto original para cada árbol. Por último cada árbol del “bosque” arroja una predicción, siendo el valor promedio como la predicción final del modelo (Parmar et al., 2019).

Los parámetros para random forest que se evaluarán de la librería de scikit learn son:

- Numero de árboles: “n\_estimators” especifica cuántos árboles de decisión se deben construir en el bosque, una cantidad de arboles mayor mejoran el rendimiento pero aumentan el costo computacional.
- Máxima profundidad del árbol: “max\_depth” controla la profundidad máxima de cada árbol de decisión en el bosque. Limitar la profundidad ayuda a controlar el sobreajuste.
- Número mínimo de muestras para dividir un nodo: “min\_samples\_split” especifica el número mínimo de muestras requeridas para dividir un nodo interno.
- Número mínimo de muestras en cada hoja: “min\_samples\_leaf” especifica el número mínimo de muestras que deben estar en cada hoja del árbol. Esto ayuda a controlar el tamaño mínimo de las hojas del árbol.
- Número máximo de características a considerar para cada división: “max\_features” determina el número máximo de características que se considerarán para realizar una división en cada nodo. Puede ser útil para reducir la correlación entre árboles y, por lo tanto, mejorar la generalización.

“Random Forest” es conocido por ser bueno para trabajar con conjuntos de datos grandes y complejos (Parmar et al., 2019).

## Regresión logística

Es un algoritmo utilizado principalmente para clasificación binaria, predice la probabilidad de que una observación sea perteneciente a una de dos posibles clases los parámetros de regresión logística son:

- Número de iteraciones máximas: durante el entrenamiento, “max\_iter” especifica el numero máximo de iteraciones antes de converger en una solución.

- Tolerancia mínima entre iteraciones: “tol” es un valor que especifica el valor mínimo aceptable entre iteraciones para detener el proceso de entrenamiento.
- Penalidad en la regularización: “penalty” es un parámetro que penaliza grandes valores de los coeficientes, existen diferentes métodos de regularización, L1, L2, Elastic Net y sin penalidad.
- Inverso de la fuerza de regularización: “C” debe ser un valor positivo. Al igual que en las máquinas de soporte vectorial, valores pequeños especifican una regularización más fuerte.

La característica más fuerte de la regresión logística es que puede predecir una variable categórica binaria, ideal en el contexto médico.

## Máquina de soporte vectorial

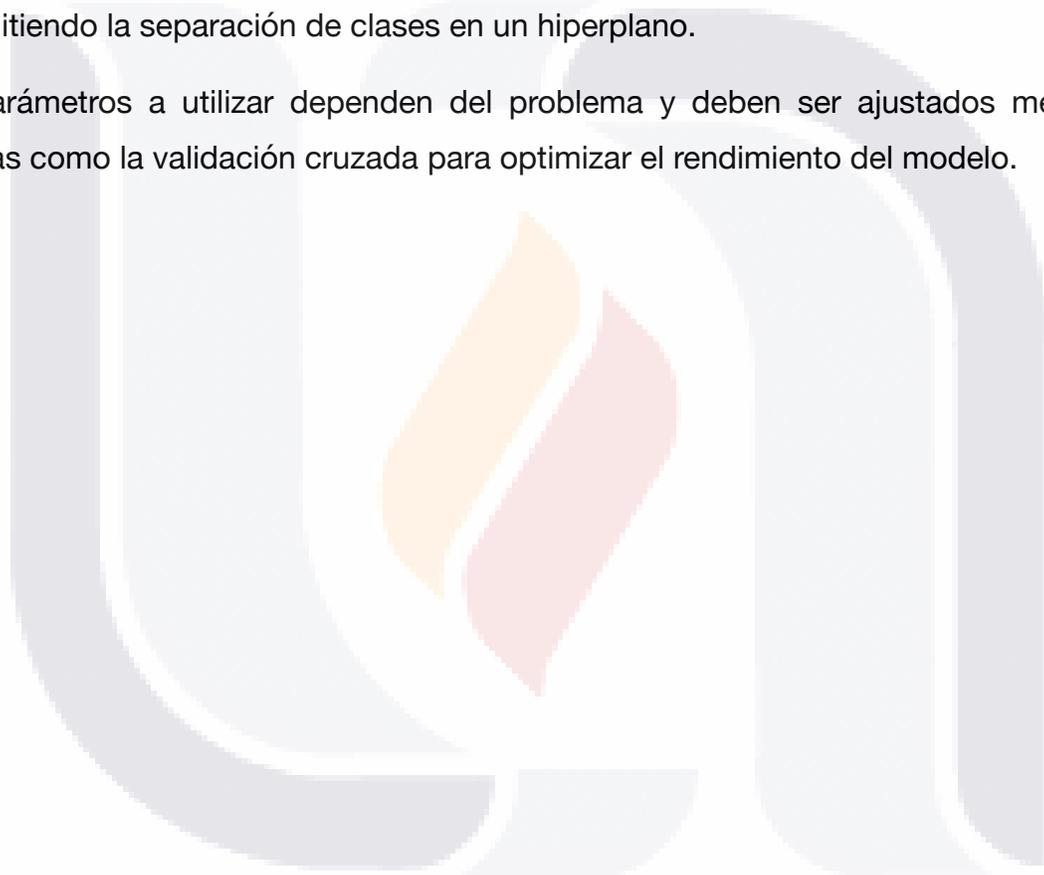
Son algoritmos que buscan encontrar el hiperplano óptimo que separa las clases en un espacio de características, este hiperplano maximiza el margen entre los puntos de datos más cercanos de las clases opuestas, estos son llamados vectores de soporte. Principalmente las SVM encuentran el hiperplano que divida a los datos en dos clases con el máximo margen, en un espacio bidimensional este hiperplano es una línea, en un espacio tridimensional es un plano, y en espacios de mayor dimensión es un hiperplano (Suthaharan, 2016).

El objetivo es encontrar el hiperplano que maximiza la distancia entre los puntos más cercanos de ambas clases. Esta maximización reduce el riesgo de clasificación errónea y mejora la capacidad del modelo para generalizar en nuevos datos. En problemas no lineales se utiliza el truco del kernel, el cual consiste en proyectar los datos a un espacio de mayor dimensión, donde estos se pueden dividir en planos lineales, entre los kernels mas comunes contamos con: linear, polinómico, función de base radial o gaussiana y sigmoide (Suthaharan, 2016).

El manejo de la regularización y el costo del error es un método que nos ayuda a tratar con datos que no son perfectamente separables, permitiendo una cierta cantidad de errores a través de la regularización (Suthaharan, 2016), los posibles parámetros son:

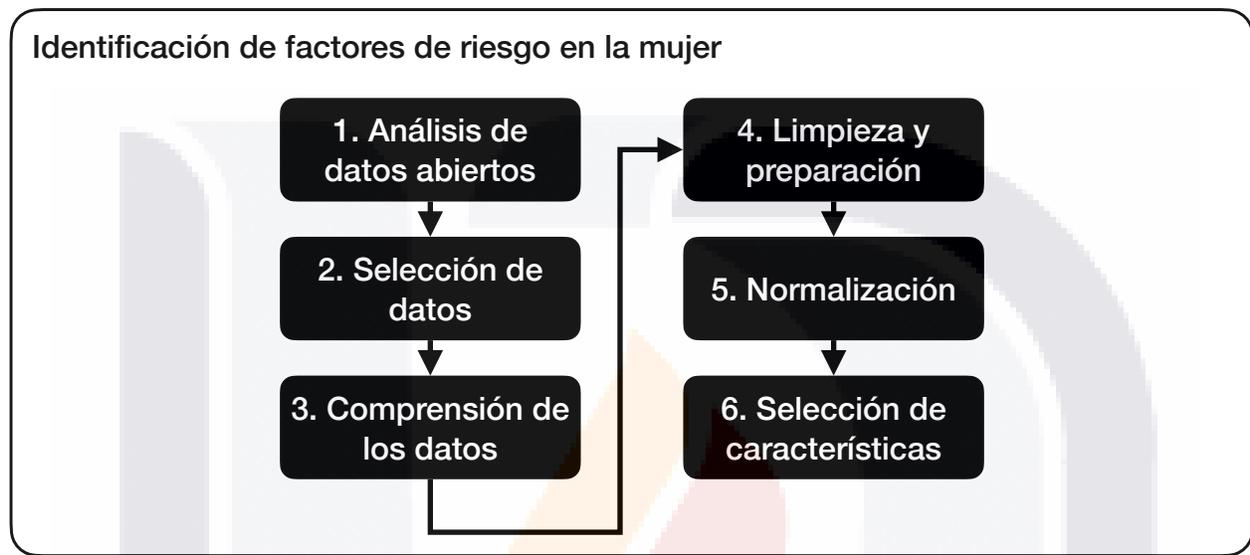
- Parámetro de regularización: “C” controla la compensación entre el ajuste correcto de los datos de entrenamiento y la maximización del margen, un valor grande permite un margen más grande, ayudando a prevenir el sobre ajuste pero puede ocasionar un mayor número de clasificaciones incorrectas.
- Estimación de probabilidades: “probability” habilita la estimación de probabilidades en la predicción.
- Tipo de kernel: “kernel” define la función de kernel para proyectar los datos, permitiendo la separación de clases en un hiperplano.

Los parámetros a utilizar dependen del problema y deben ser ajustados mediante técnicas como la validación cruzada para optimizar el rendimiento del modelo.



## VII. Metodología

Para encontrar los factores de riesgo en las mujeres se plantea la siguiente metodología, la cuál nos ayudará a seleccionar aquellas variables que logren determinar el riesgo de infarto y con esta selección generar una herramienta para el apoyo en la toma de decisiones (Figura 8).

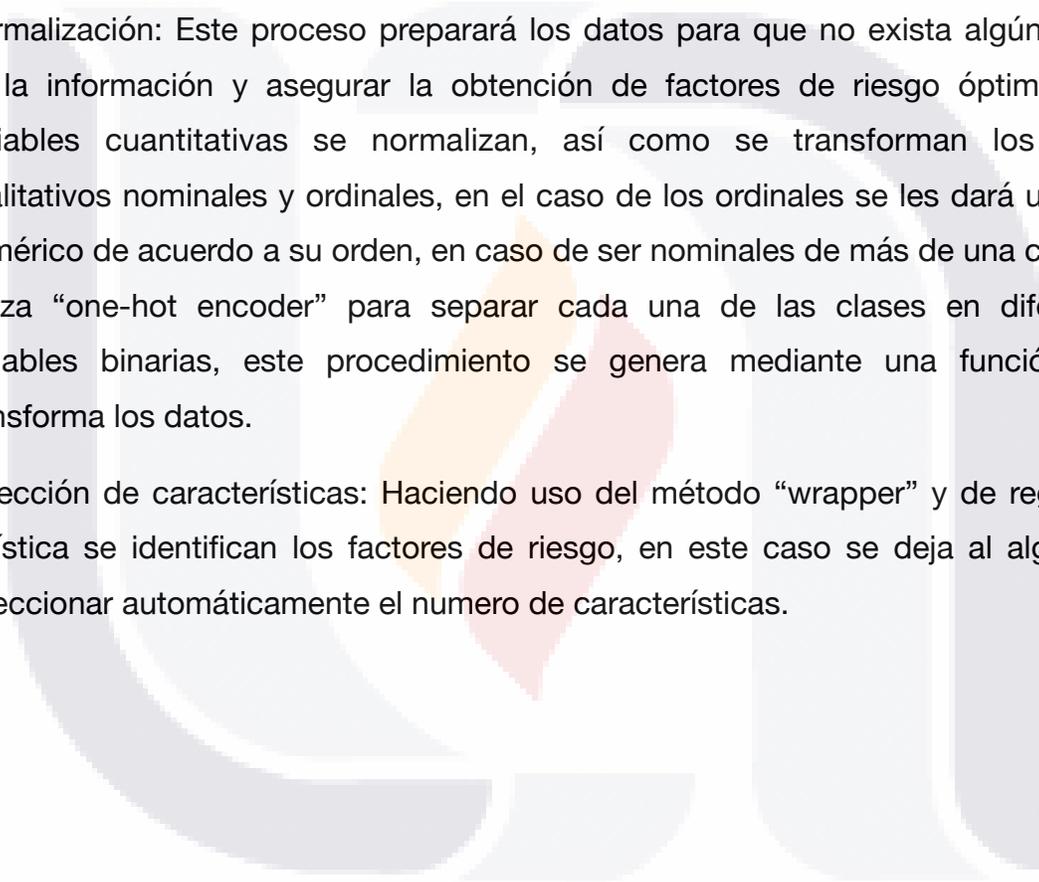


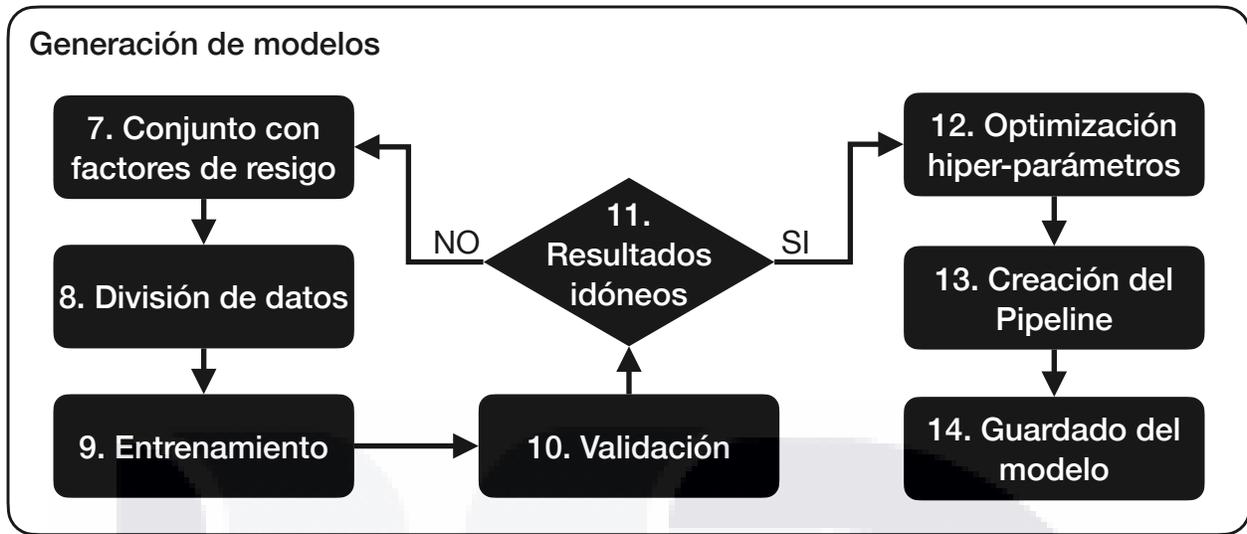
**FIG 8.** Metodología: Identificación de factores de riesgo de ECV en la mujer.

En la figura 8 se pueden observar los pasos a seguir para la obtención de los factores de riesgo de los distintos conjuntos de datos a utilizar, esta primera fase de la metodología tiene como objetivo obtener los factores de riesgo específicos en la mujer.

Para cada una de las diferentes bases de datos se realiza el siguiente procedimiento:

1. Análisis de datos abiertos: Se contextualiza de dónde proviene la información y cuál es el contexto de la variable dependiente, esto nos servirá para conocer el dominio en el que podría ser utilizado el modelo resultante.
2. Selección de datos: se seleccionan únicamente las mujeres del conjunto de datos, además se eliminan las filas que contengan datos nulos, esto debido a que al ser un problema médico, es importante tener datos auténticos y no sintéticos.

- TESIS TESIS TESIS TESIS TESIS
3. Comprensión de los datos: Verificar el dominio de los datos y que estos correspondan con la literatura médica, enseguida validar la unidad de medida y de ser necesario homologarla a una unidad de medida estándar.
  4. Limpieza y preparación: Esta etapa contempla la homologación y normalización de las variables: las columnas se re-nombran de forma que sean comprensibles y los valores que contienen se estandarizan, el estado actual del los datos se guarda, además, se balancea la variable dependiente del conjunto de datos.
  5. Normalización: Este proceso preparará los datos para que no exista algún sesgo en la información y asegurar la obtención de factores de riesgo óptimos. las variables cuantitativas se normalizan, así como se transforman los datos cualitativos nominales y ordinales, en el caso de los ordinales se les dará un valor numérico de acuerdo a su orden, en caso de ser nominales de más de una clase se utiliza “one-hot encoder” para separar cada una de las clases en diferentes variables binarias, este procedimiento se genera mediante una función que transforma los datos.
  6. Selección de características: Haciendo uso del método “wrapper” y de regresión logística se identifican los factores de riesgo, en este caso se deja al algoritmo seleccionar automáticamente el numero de características.
- 

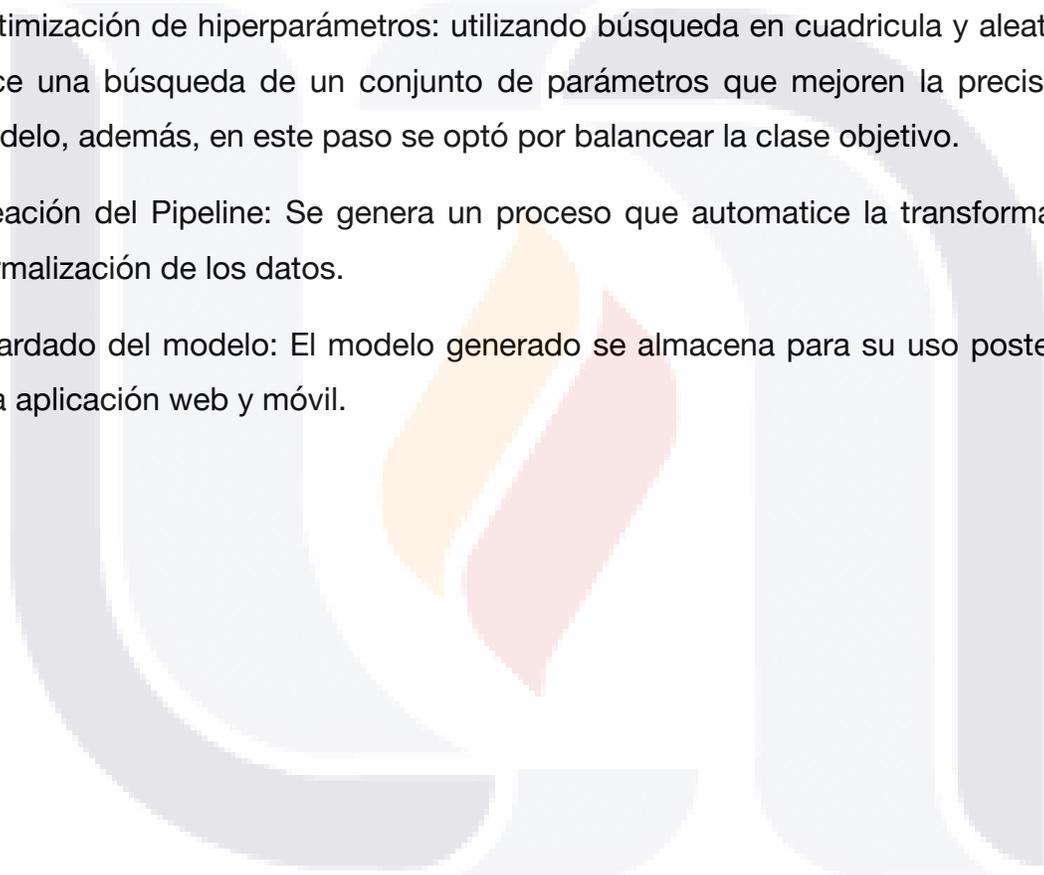


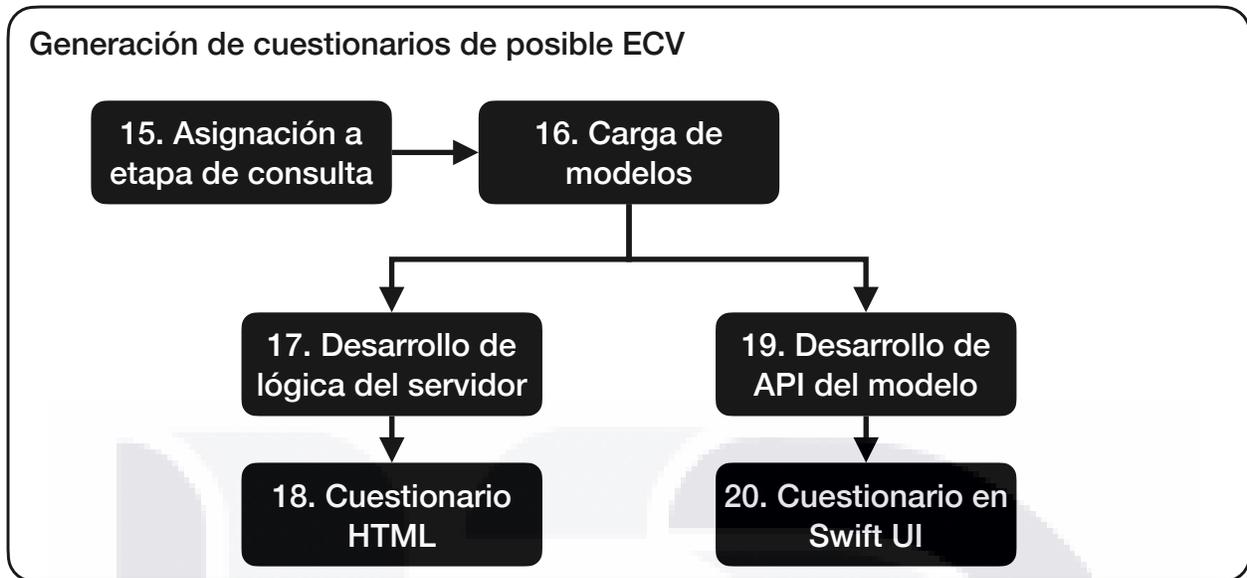
**FIG 9.** Metodología: Generación de modelos.

En la segunda etapa de la metodología, ya habiendo seleccionado los factores de riesgo para cada conjunto de datos, se procede a la generación de modelos (Figura 9).

7. Conjunto con factores de riesgo: este conjunto de datos es el set de datos original, esto para tener la posibilidad de más adelante poder transformar los datos dentro de un Pipeline, este set de datos es el que almacenamos en el paso 4 (Figura 8). Leemos este set de datos pero únicamente consideramos los factores de riesgo seleccionados.
8. División de datos: se realiza una división de datos de forma aleatoria con un porcentaje de separación para datos de entrenamiento de 80% y el restante para validación.
9. Entrenamiento: Al entrenar el modelo se utiliza la función de normalización generada en el paso 5 (Figura 8). Se envía una copia de los datos originales para la transformación esto nos ayudará a evitar transformar los datos que posteriormente se utilizaran para el modelo final, en este paso se entrenan 3 modelos utilizando Bosque aleatorio, regresión logística y maquina de soporte vectorial.

10. Validación: Se evalúan los 3 modelos utilizando una copia del conjunto de datos de validación procesada por la función de normalización generada en el paso 5 (Figura 8). Los resultados de la precisión se guardan para ser evaluados posteriormente.
11. Resultados idóneos: se comparan los resultados de cada modelo, si alguno de los modelos consigue repetidamente mejores resultados a los demás modelos, este será seleccionado, y los modelos que no logren una buena evaluación se descartaran para el conjunto de datos actual.
12. Optimización de hiperparámetros: utilizando búsqueda en cuadrícula y aleatoria se hace una búsqueda de un conjunto de parámetros que mejoren la precisión del modelo, además, en este paso se optó por balancear la clase objetivo.
13. Creación del Pipeline: Se genera un proceso que automatice la transformación y normalización de los datos.
14. Guardado del modelo: El modelo generado se almacena para su uso posterior en una aplicación web y móvil.





**FIG 10.** Metodología: Cuestionarios de posible ECV.

En la tercer y última etapa se concibe una herramienta mediante la cual sea posible el uso de los modelos conforme al contexto de la variable dependiente (Figura 10).

15. Asignación a etapa de consulta: Dependiendo de los factores de riesgo seleccionados para el conjunto de datos y de la variable dependiente se genera una recomendación que se adecue al proceso de consulta del paciente.

16. Carga de modelos: Se configura un entorno web mediante el uso de la librería de python “Flask”, se genera un proyecto al cual se le cargan los modelos generados.

17. Desarrollo de lógica del servidor Codificación de la lógica backend del servidor, el cual utilizara el método post para proporcionar una respuesta que contenga la clasificación del modelo.

18. Cuestionario HTML: Haciendo uso del framework de Bootstrap y HTML se codifica un cuestionario sencillo, el cual está disponible en la sección de resultados.

19. Desarrollo de API del modelo: Bajo una ruta especifica se añade la lógica necesaria para responder con la clasificación del modelo.

20. Cuestionario en SwiftUI: haciendo uso de la tecnología de SwiftUI se codifica un formulario que pueda hacer consultas al servidor.

## Material

Para la realización de este estudio, se seleccionaron cinco conjuntos de datos abiertos de la plataforma Kaggle. Al elegir estos conjuntos, se tuvo en cuenta que debían incluir registros de mujeres.

### Conjunto de datos de Framingham

En la ciudad de Framingham, Massachusetts, bajo la supervisión del “National Heart, Lung, and Blood Institute” (NHLBI) se lleva a cabo un estudio longitudinal cuyo objetivo es valorar la evolución de distintas ECV (National Heart, 2023). En el 2018, como resultado de esta investigación, se publicó un conjunto de datos que cuenta con 4,000 registros, 15 atributos y una variable dependiente (Tabla 3) (Mirza, 2023).

TABLA 3. Descripción de las variables del conjunto de datos de Framingham (Mirza, 2023).

Variable	Descripción	Tipo de variable
<b>sex</b>	Sexo de la persona que pertenece al estudio.	Cualitativa Nominal
<b>age</b>	Edad de la persona que pertenece al estudio.	Cuantitativa Discreta
<b>education</b>	Último grado de estudios.	Cualitativa Ordinal
<b>cigsPerDay</b>	Cantidad de cigarrillos consumidos al día.	Cuantitativa Discreta
<b>BPMeds</b>	Consumo de medicamento para la presión arterial alta.	Cualitativa Nominal
<b>prevalentStroke</b>	Padeció de evento cerebrovascular.	Cualitativa Nominal
<b>prevalentHyp</b>	Sufre de hipertensión.	Cualitativa Nominal
<b>diabetes</b>	Sufre de diabetes.	Cualitativa Nominal
<b>totChol</b>	Niveles de colesterol en sangre mg/dL.	Cuantitativa discreta
<b>sysBP</b>	Presión arterial sistólica.	Cuantitativa discreta
<b>diaBP</b>	Presión arterial diastólica.	Cuantitativa discreta
<b>BMI</b>	Índice de masa corporal.	Cuantitativa Continua
<b>heartRate</b>	Ritmo cardiaco de la persona que pertenece al estudio.	Cuantitativa discreta
<b>Glucose</b>	Niveles de glucosa en sangre mg/dL.	Cuantitativa discreta
<b>Ten Year CHD</b>	Variable dependiente de posible enfermedad coronaria en los próximos 10 años	Cualitativa Nominal

## Conjunto de datos combinado

Los autores utilizaron cinco conjuntos de datos con población proveniente de Estados Unidos: Cleveland y Long Beach, Hungría, Suiza y un conjunto de datos “Statlog (Heart)” proveniente del repositorio “UC Irvine” (Fedesoriano, 2021).

Los autores encontraron un total de 11 atributos en común, dieron un tratamiento a los datos eliminando duplicados, obteniendo un total de 918 registros (Fedesoriano, 2021).

Esta base de datos cuenta con un total de 12 características, en ellas una variable dependiente, la descripción de cada atributo los encontramos en la Tabla 4.

TABLA 4. Descripción de las variables del conjunto de datos combinados (Fedesoriano, 2021).

Variable	Descripción	Tipo de variable
<b>Age</b>	Edad de la persona que pertenece al estudio.	Cuantitativa Discreta
<b>Sex</b>	Sexo de la persona que pertenece al estudio.	Cualitativa Nominal
<b>ChestPainType</b>	Tipo de dolor torácico	Cualitativa Ordinal
<b>RestingBP</b>	mm Hg	Cuantitativa Discreta
<b>Cholesterol</b>	mm/dL	Cualitativa Nominal
<b>FastingBS</b>	Azúcar en sangre en ayunas, si es mayor a 120 mg/dL	Cualitativa Nominal
<b>RestingECG</b>	Interpretación del electrocardiograma durante el reposo.	Cualitativa Nominal
<b>Oldpeak</b>	Depresión del segmento ST.	Cuantitativa Continua
<b>MaxHR</b>	Ritmo cardíaco durante la prueba de esfuerzo.	Cuantitativa Discreta
<b>ExerciseAngina</b>	Angina inducida por prueba de esfuerzo.	Cuantitativa Discreta
<b>ST_Slope</b>	Segmento ST después de la prueba de esfuerzo.	Cualitativa Nominal
<b>HeartDisease</b>	Variable dependiente de ECV diagnosticado en el paciente.	Cualitativa Nominal

## Conjunto de datos CDC 2020

Este conjunto de datos proviene del sistema de factores de riesgo del comportamiento “Behavioral Risk Factor Surveillance System” del centro para el control y prevención de enfermedades en Estados Unidos, o CDC por sus siglas en inglés; fue publicado en Kaggle por el usuario Yamil Pytlak (Pytlak, 2020). Cuenta con 319,794 observaciones, 16 variables independientes y una dependiente (Tabla 5).

TABLA 5. Descripción de las variables del conjunto de datos del 2020 (Pytlak, 2020).

Variable	Descripción	Tipo de variable
<b>HeartDisease</b>	Variable dependiente de ECV.	Cuantitativa Discreta
<b>BMI</b>	Índice de masa corporal.	Cualitativa Nominal
<b>Smoking</b>	Fumador actual.	Cualitativa Ordinal
<b>AlcoholDrinking</b>	Consumo frecuente de alcohol.	Cuantitativa Discreta
<b>Stroke</b>	Sufrió de accidente cerebrovascular.	Cualitativa Nominal
<b>PhysicalHealth</b>	Estado de salud física en 5 puntos.	Cualitativa Nominal
<b>MentalHealth</b>	Estado de salud mental en 5 puntos.	Cualitativa Nominal
<b>DiffWalking</b>	Dificultad al caminar.	Cualitativa Nominal
<b>AgeCategory</b>	Categoría de edad.	Cuantitativa Discreta
<b>Race</b>	Grupo étnico de la persona.	Cuantitativa Discreta
<b>Diabetic</b>	Clase de diabetes, indica si tuvo diabetes gestacional.	Cuantitativa Discreta
<b>PhysicalActivity</b>	Actividad física los últimos 30 días.	Cuantitativa Discreta
<b>GenHealth</b>	Salud en general los últimos 30 días.	Cualitativa Ordinal
<b>SleepTime</b>	Tiempo de sueño en promedio al día.	Cuantitativa Discreta
<b>Asthma</b>	Si padece asma.	Cualitativa Nominal
<b>KidneyDisease</b>	Si padece enfermedad de los riñones.	Cualitativa Nominal
<b>SkinCancer</b>	Si padece cáncer de piel.	Cualitativa Nominal

## Conjunto de datos CDC 2022

Este conjunto de datos también proviene del sistema de factores de riesgo del comportamiento “Behavioral Risk Factor Surveillance System” del CDC en Estados Unidos, el usuario Yamil Pytlak de Kaggle publico los resultados de la encuesta llevada a cabo en el 2022 (Pytlak, 2022). Cuenta 246,020 registros, 39 atributos y una variable dependiente (Tabla 6).

TABLA 6. Descripción de las variables del conjunto de datos del 2022 (Pytlak, 2022).

Variable	Descripción	Tipo de variable
<b>State</b>	Estado o región en la que reside la persona.	Cuantitativa Discreta
<b>Sex</b>	Sexo biológico de la persona.	Cualitativa Nominal
<b>GeneralHealth</b>	Percepción general de la salud de una persona.	Cualitativa Ordinal
<b>PhysicalHealthDays</b>	Número de días en el último mes que una persona ha experimentado problemas de salud física.	Cuantitativa discreta
<b>MentalHealthDays</b>	Número de días en el último mes que una persona ha experimentado problemas de salud mental o emocional.	Cualitativa Nominal
<b>LastCheckupTime</b>	Última vez que la persona se realizó un chequeo médico de rutina.	Cualitativa Nominal
<b>PhysicalActivities</b>	Participación en actividades físicas fuera del trabajo en los últimos 30 días.	Cualitativa Nominal
<b>SleepHours</b>	Promedio de horas de sueño por día en un período de 24 horas.	Cualitativa Nominal
<b>RemovedTeeth</b>	Número de dientes removidos o perdidos por la persona.	Cuantitativa discreta
<b>HadAngina</b>	Indica si una persona ha sido diagnosticada con angina de pecho.	Cuantitativa discreta
<b>HadStroke</b>	Indica si una persona ha sufrido un accidente cerebrovascular.	Cuantitativa Continua
<b>HadAsthma</b>	Indica si una persona ha sido diagnosticada con asma.	Cualitativa Nominal
<b>HadSkinCancer</b>	Indica si una persona ha sido diagnosticada con cáncer de piel.	Cualitativa Nominal
<b>HadCOPD</b>	Indica si una persona ha sido diagnosticada con Enfermedad Pulmonar Obstructiva Crónica (EPOC).	Cualitativa Nominal
<b>HadDepressiveDisorder</b>	Indica si una persona ha sido diagnosticada con un trastorno depresivo.	Cualitativa Nominal
<b>HadKidneyDisease</b>	Indica si una persona ha sido diagnosticada con enfermedad renal crónica.	Cualitativa Nominal
<b>HadArthritis</b>	Indica si una persona ha sido diagnosticada con artritis.	Cualitativa Nominal

Variable	Descripción	Tipo de variable
<b>HadDiabetes</b>	Indica si una persona ha sido diagnosticada con diabetes.	Cualitativa Nominal
<b>DeafOrHardOfHearing</b>	Indica si una persona tiene pérdida de audición parcial o completa.	Cualitativa Nominal
<b>BlindOrVisionDifficulty</b>	Indica si una persona tiene problemas de visión severos o es ciega.	Cualitativa Nominal
<b>DifficultyConcentrating</b>	Indica si una persona tiene dificultades para concentrarse, recordar o tomar decisiones debido a una condición física, mental o emocional.	Cualitativa Nominal
<b>DifficultyWalking</b>	Indica si una persona tiene dificultades para caminar o moverse.	Cualitativa Nominal
<b>DifficultyDressingBathing</b>	Indica si una persona tiene dificultades para vestirse o bañarse por sí misma.	Cualitativa Nominal
<b>DifficultyErrands</b>	Indica si una persona tiene dificultades para realizar actividades solos.	Cualitativa Nominal
<b>SmokerStatus</b>	Estado de consumo de tabaco de una persona.	Cualitativa Nominal
<b>ECigaretteUsage</b>	Uso de cigarrillos electrónicos por parte de una persona.	Cualitativa Nominal
<b>ChestScan</b>	Indica si una persona ha tenido un escáner de tórax, como una radiografía o tomografía computarizada, para examinar los pulmones.	Cualitativa Nominal
<b>RaceEthnicityCategory</b>	Clasificación de la persona en una categoría de raza o etnia.	Cualitativa Nominal
<b>AgeCategory</b>	Clasificación de la edad de una persona en diferentes rangos.	Cualitativa Ordinal
<b>HeightInMeters</b>	Altura de la persona medida en metros.	Cuantitativa Continua
<b>WeightInKilograms</b>	Peso de la persona medido en kilogramos.	Cuantitativa Continua
<b>BMI</b>	Índice de Masa Corporal (IMC), que mide la relación entre el peso y la altura de una persona.	Cuantitativa Discreta
<b>AlcoholDrinkers</b>	Indica si una persona consume alcohol regularmente.	Cualitativa Nominal
<b>HIVTesting</b>	Indica si una persona se ha realizado una prueba de VIH.	Cualitativa Nominal
<b>FluVaxLast12</b>	Indica si una persona ha recibido la vacuna contra la gripe en los últimos 12 meses.	Cualitativa Nominal
<b>PneumoVaxEver</b>	Indica si una persona ha recibido alguna vez la vacuna neumocócica.	Cualitativa Nominal
<b>TetanusLast10Tdap</b>	Indica si una persona ha recibido la vacuna contra el tétanos en los últimos 10 años.	Cualitativa Nominal
<b>HighRiskLastYear</b>	Indica si una persona ha tenido comportamientos de alto riesgo en el último año.	Cualitativa Nominal
<b>CovidPos</b>	Indica si una persona ha dado positivo en una prueba de COVID-19.	Cualitativa Nominal
<b>HadHeartAttack</b>	Variable dependiente de ataque cardíaco.	Cualitativa Nominal

## Conjunto de datos Punjab

Este conjunto de datos incluye información clínica recopilada de pacientes atendidos en el Instituto de Cardiología de Faisalabad, ubicado en Punjab, Pakistán, durante el período de Abril a Diciembre del 2015. Este compendio de datos está compuesto por 299 registros individuales. Cada registro abarca 11 atributos clínicos, además se incluye una característica informativa específica del tiempo de seguimiento del paciente, junto con una variable dependiente que permite evaluar y predecir los resultados de salud cardiovascular (Tabla 7). Esta colección de datos contiene pacientes en edades avanzadas (Chicco & Jurman, 2020).

TABLA 7. Descripción de las variables del conjunto de datos de Punjab (Chicco & Jurman, 2020).

Variable	Descripción	Tipo de variable
<b>age</b>	La edad del paciente	Cuantitativa Discreta
<b>anemia</b>	Si el paciente tiene anemia	Cualitativa Nominal
<b>creatinine_phosphokinase</b>	Levels of CPK enzyme in blood, mcg/L	Cualitativa Ordinal
<b>diabetes</b>	Si el paciente tiene diabetes	Cuantitativa discreta
<b>ejection_fraction</b>	Porcentaje de sangre que sale del corazón en cada contracción	Cualitativa Nominal
<b>high_blood_pressure</b>	Si el paciente tiene presión arterial alta	Cualitativa Nominal
<b>platelets</b>	Conteo de plaquetas en sangre.	Cualitativa Nominal
<b>serum_creatinine</b>	Nivel de creatinina sérica en la sangre mg/dl	Cualitativa Nominal
<b>serum_sodium</b>	Nivel de sodio sérico en la sangre mEq/L	Cuantitativa discreta
<b>sex</b>	Sexo del paciente	Cuantitativa discreta
<b>smoking</b>	Si el paciente fuma	Cuantitativa Continua
<b>time</b>	El período de seguimiento en días	Cuantitativa discreta
<b>DEATH_EVENT</b>	Si el paciente falleció durante el período de seguimiento	Cualitativa Nominal

## VIII.Resultados

Después de realizar los experimentos se obtuvieron los siguientes factores de riesgo en mujeres para cada uno de los conjuntos de datos:

- Para el conjunto de datos de Framingham (National Heart, 2023):
  - Sexo, nivel educativo, si es fumador actual, cigarrillos al día, sufrió evento cerebrovascular, sufre hipertensión, tiene hipertensión medicada, sufre diabetes y el IMC actual son factores de riesgo de contraer una ECV en los próximos 10 años.
- Para el conjunto de datos de distintos estudios (Fedesoriano, 2021):
  - Edad, tipo de angina, glucosa en sangre, interpretación de EKG en descanso, depresión del segmento ST, Angina inducida por prueba de esfuerzo y el comportamiento del segmento ST después de prueba de esfuerzo.
- El conjunto de datos del CDC 2020 (Pytlak, 2020):
  - Edad, padecer asma, evento cerebrovascular, dificultad al caminar, fumador actual, salud en general, diabetes gestacional y enfermedad de los riñones.
- Del conjunto de datos del CDC 2022 (Pytlak, 2022):
  - Edad, cantidad de dientes removidos, salud en general, tiempo de sueño, raza, último chequeo médico, angina de pecho, estatura, padecer VIH, evento cerebrovascular, padecer cáncer de piel, el alcoholismo, las enfermedades de los riñones, la dificultad al caminar, el uso de cigarrillo, no haber tenido vacunas de neumonía y de la gripe y padecer diabetes.
- Del conjunto de datos de Punjab Pakistan (Chicco & Jurman, 2020):
  - Presión arterial alta, niveles de creatinina sérica, niveles de suero de sodio, si es fumador actual, si padece anemia y si padece diabetes son factores de riesgo que aumentan la mortalidad al padecer un IAM.

Enseguida se muestran los resultados para las siguientes etapas de la metodología, primero veremos los resultados de los diferentes modelos.

### Generación de modelos

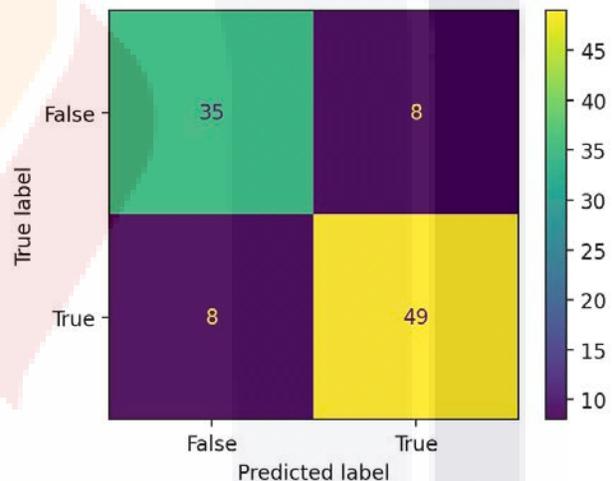
Para la generación de todos los modelos se utilizó el siguiente hardware y software:

- iPad Pro cuarta generación, 8 Núcleos, 8 GB de memoria.
- Juno 3.1.0 (9191) (Jupiter Notebooks).
- Python 3.10.

### Conjunto de datos de Framingham

Se obtuvieron un total de 2,034 mujeres para este análisis, el modelo clasificador utilizado esta basado en Support Vector Machine. Los resultados obtenidos en el conjunto de entrenamiento y evaluación son los siguientes.

- Entrenamiento:
  - Puntuación F1: 67%
- Evaluación:
  - Exactitud (Accuracy): 84%
  - Precisión (Precision): 85.9%
  - Recall: 85.96%
  - Puntuación F1: 85.96%
  - El Kappa de Cohen: 67.36%
  - Área Bajo la Curva (AUC): 86.98%



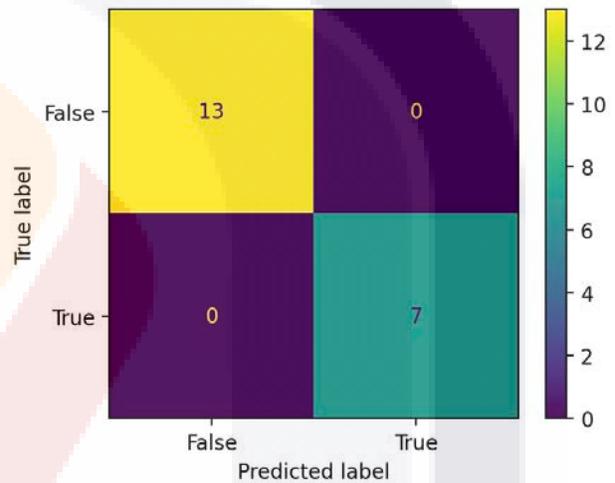
**FIG 11.** Matriz de confusión, modelo 1.

La figura 11 nos muestra la matriz de confusión para la evaluación de este modelo, obteniendo una precisión del 81% en la clase negativa y un 85% en la clase positiva.

### Conjunto de datos combinado

Se obtuvieron un total de 193 registros de mujeres del conjunto de datos combinado (Fedesoriano, 2021). Estos datos fueron utilizados para generar el modelo clasificador utilizando Regresión Logística (Logistic Regression). A continuación, se presentan los resultados obtenidos tanto en el conjunto de entrenamiento como en el conjunto de evaluación.

- Entrenamiento:
  - Puntuación F1: 80%
- Evaluación:
  - Exactitud (Accuracy): 100%
  - Precisión (Precision): 100%
  - Recall: 100%
  - Puntuación F1: 100%
  - El Kappa de Cohen: 100%
  - Área Bajo la Curva (AUC): 100%



**FIG 12.** Matriz de confusión, modelo 2.

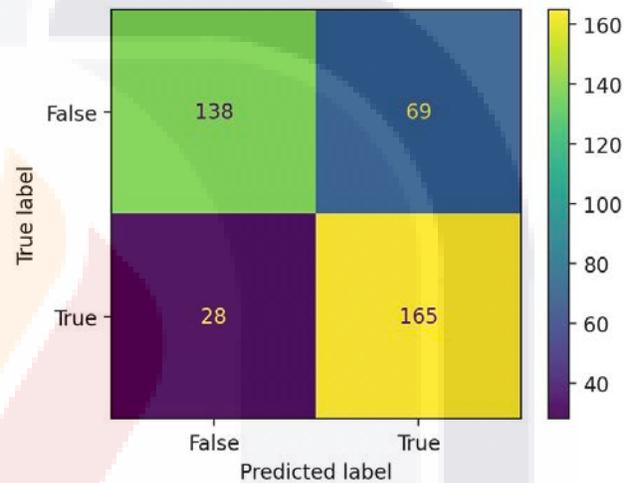
La figura 12 nos muestra la matriz de confusión para la evaluación de este modelo, obteniendo una precisión del 100% en la clase negativa y un 100% en la clase positiva.

### Conjunto de datos CDC 2020

Se obtuvieron un total de 142,460 registros de mujeres del conjunto de datos del CDC de la encuesta realizada en el 2020 (Pytlak, 2020).

En este análisis, se utilizó la Regresión Logística (Logistic Regression) como modelo clasificador. A continuación, se presentan los resultados obtenidos en los conjuntos de entrenamiento y evaluación.

- Entrenamiento:
  - Puntuación F1: 79%
- Evaluación:
  - Exactitud (Accuracy): 75.75%
  - Precisión (Precision): 70.51%
  - Recall: 85.49%
  - Puntuación F1: 77.28%
  - El Kappa de Cohen: 51.78%
  - Área Bajo la Curva (AUC): 82.52%



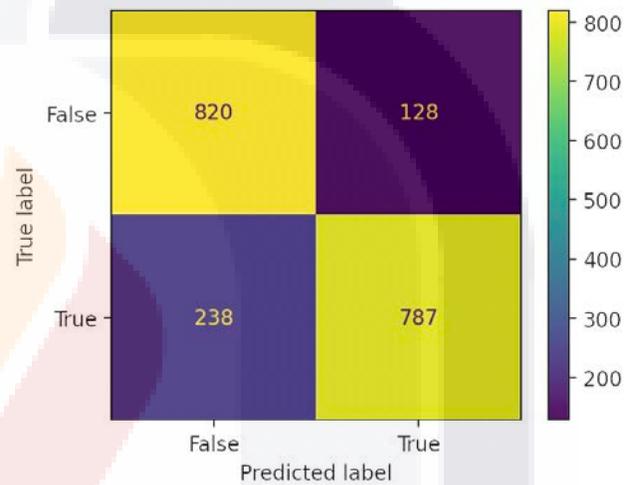
**FIG 13.** Matriz de confusión, modelo 3.

La figura 13 nos muestra la matriz de confusión para la evaluación de este modelo, obteniendo una precisión del 66% en la clase negativa y un 85% en la clase positiva.

### Conjunto de datos CDC 2022

Para el conjunto de datos del CDC 2022 (Pytlak, 2022), se obtuvieron un total de 127,811 registros de mujeres. Se utilizó la Regresión Logística (Logistic Regression) como modelo clasificador. A continuación, se presentan los resultados obtenidos tanto en el conjunto de entrenamiento como en el conjunto de evaluación.

- Entrenamiento:
  - Puntuación F1: 77%
- Evaluación:
  - Exactitud (Accuracy): 81.44%
  - Precisión (Precision): 86.01%
  - Recall: 76.78%
  - Puntuación F1: 81.13%
  - El Kappa de Cohen: 63%
  - Área Bajo la Curva (AUC): 89.40%



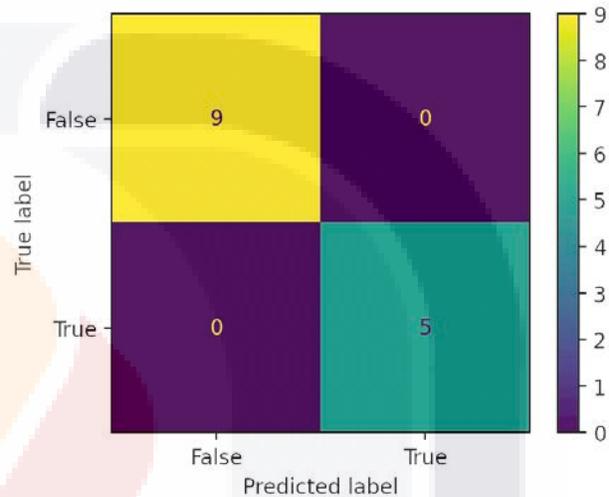
**FIG 14.** Matriz de confusión, modelo 4.

La figura 14 nos muestra la matriz de confusión para la evaluación de este modelo, obteniendo una precisión del 86% en la clase negativa y un 76% en la clase positiva.

### Conjunto de datos Punjab, Pakistan

Se obtuvieron un total de 142,460 registros de mujeres del conjunto de datos de Punjab, Pakistan (Chicco & Jurman, 2020), para este modelo se utilizó la Regresión Logística (Logistic Regression) como modelo clasificador. A continuación, se detallan los resultados obtenidos en los conjuntos de entrenamiento y evaluación:

- Entrenamiento:
  - Puntuación F1: 89%
- Evaluación:
  - Exactitud (Accuracy): 100%
  - Precisión (Precision): 100%
  - Recall: 100%
  - Puntuación F1: 100%
  - El Kappa de Cohen: 100%
  - Área Bajo la Curva (AUC): 100%



**FIG 15.** Matriz de confusión, modelo 5.

La figura 15 nos muestra la matriz de confusión para la evaluación de este modelo, obteniendo una precisión del 100% en la clase negativa y un 100% en la clase positiva.

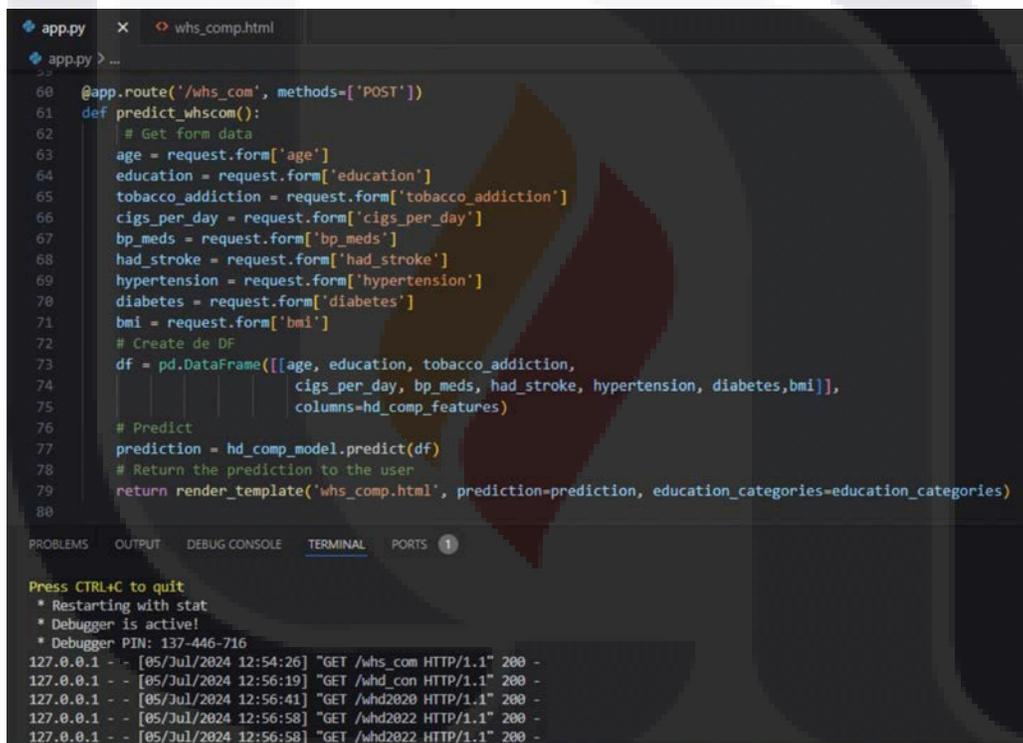
En base a este modelo se realizó un artículo dentro del cual se mencionan los pasos seguidos y los resultados obtenidos haciendo uso de la herramienta de generación de modelos de aprendizaje automático Create ML de Apple (Create ML | Apple Developer Documentation, 2023), los resultados obtenidos con esta herramienta no fueron tan elevados como los obtenidos en esta iteración obteniendo un (Anexo B).

## Servidor

El sistema operativo seleccionado para el servidor es Windows®. A continuación se detallan sus especificaciones técnicas y del software utilizado:

- HP Pavilion 2016, Intel® i5, 4 Núcleos, 8 GB de memoria.
- Flask 3.0.1.
- Python 3.10.

En la figura 16 se aprecia el proyecto desarrollado haciendo uso del marco de trabajo Flask en Python, el desarrollo de esta aplicación fue utilizando Visual Studio Code®.



```
app.py x whs_comp.html
app.py > ...
60 @app.route('/whs_com', methods=['POST'])
61 def predict_whscom():
62     # Get form data
63     age = request.form['age']
64     education = request.form['education']
65     tobacco_addiction = request.form['tobacco_addiction']
66     cigs_per_day = request.form['cigs_per_day']
67     bp_meds = request.form['bp_meds']
68     had_stroke = request.form['had_stroke']
69     hypertension = request.form['hypertension']
70     diabetes = request.form['diabetes']
71     bmi = request.form['bmi']
72     # Create de DF
73     df = pd.DataFrame([[age, education, tobacco_addiction,
74                       cigs_per_day, bp_meds, had_stroke, hypertension, diabetes, bmi]],
75                      columns=hd_comp_features)
76     # Predict
77     prediction = hd_comp_model.predict(df)
78     # Return the prediction to the user
79     return render_template('whs_comp.html', prediction=prediction, education_categories=education_categories)
80

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS 1

Press CTRL+C to quit
* Restarting with stat
* Debugger is active!
* Debugger PIN: 137-446-716
127.0.0.1 - - [05/Jul/2024 12:54:26] "GET /whs_com HTTP/1.1" 200 -
127.0.0.1 - - [05/Jul/2024 12:56:19] "GET /whd_con HTTP/1.1" 200 -
127.0.0.1 - - [05/Jul/2024 12:56:41] "GET /whd2020 HTTP/1.1" 200 -
127.0.0.1 - - [05/Jul/2024 12:56:58] "GET /whd2022 HTTP/1.1" 200 -
127.0.0.1 - - [05/Jul/2024 12:56:58] "GET /whd2022 HTTP/1.1" 200 -
```

FIG 16. Captura de pantalla del proyecto del servidor en VS Code.

El código pertenece al archivo principal utilizado para ejecutar el proyecto (Fig. 16), en primer lugar tenemos un decorador “@app.route” en la parte superior de la función “predict\_whs\_com” el cual permite ejecutar esta función cuando el servidor reciba una petición POST en la ruta “/whs\_com” correspondiente al primer cuestionario, el cual esta descrito más adelante en la sección “Cuestionario WEB”.

Dentro de esta función (Fig. 16) encontramos el código necesario para realizar la recolección de datos del formulario enviado por el usuario, cada línea extrae un valor específico del cuerpo de la petición. Al tener cada uno de los valores necesarios para el modelo se genera una variable de tipo "DataFrame" de pandas, esta estructura de datos tabular se genera teniendo el cuidado de que las columnas coincidan con las que cada modelo requiere, en este caso, las columnas declaradas en la línea 75 como "hd\_comp\_features". Enseguida, en la línea 77, se hace uso del modelo de predicción previamente entrenado (hd\_comp\_model), la variable df de tipo "DataFrame" se pasa al método "predict" del modelo para generar una predicción.

Finalmente en la línea 77 (Fig. 16), el resultado de la predicción se envía de nuevo al usuario mediante la función "render\_template" la cual utiliza un archivo HTML del directorio de trabajo del proyecto. Esta función utiliza la plantilla "whs\_comp.html", pasándole la predicción y otras variables categóricas como "education\_categories" que se utilizan para mostrar el cuestionario con los resultados al usuario.

En la parte inferior de la figura 16, la terminal muestra el proyecto en ejecución. La línea "Debugger is active!" indica que la aplicación se está ejecutando en modo de depuración para detectar y solucionar problemas en tiempo real. A continuación, se muestran las solicitudes HTTP que la aplicación ha recibido, junto con sus respectivas rutas. Entre estos, se incluyen "/whs\_com", que activa la función `predict\_whscom`, así como "/whd\_con", "/whd2020", y "/whd2022", otros endpoints manejados por la aplicación. Cada solicitud registrada incluye la dirección IP de origen 127.0.0.1, indicando una solicitud local, la fecha y hora de la solicitud, el método HTTP utilizado, GET, la ruta solicitada y el estado de la respuesta 200, indicando éxito.

Este proyecto cuenta con rutas tanto para la aplicación WEB como para la API responsable de enviar el resultado del modelo a la aplicación móvil iOS®. La lógica para los otros modelos es idéntica, variando únicamente las variables recolectadas para adaptarse al modelo específico. En el caso de las rutas consumidas por la aplicación móvil, estas terminan en /api, como por ejemplo: "/whd\_con/api".

### Cuestionario WEB

Estos cuestionarios están enfocados en conocer el riesgo de ECV, apoyo diagnóstico y conocer la mortalidad de un IAM en mujeres, el código de éste proyecto es accesible desde el repositorio de GitHub® [https://github.com/rdls21/women\\_hd\\_web](https://github.com/rdls21/women_hd_web).

### Women Heart Stroke Complement Dataset

Age:

Education:

BMI:

Tobacco Addiction:

Cigarettes Per-day:

Suffers Hypertension:

Medicated Hypertension:

Suffers Diabetes:

Suffered Stroke:

**FIG 17.** Cuestionario del riesgo de ECV en 10 años.

La figura 17 muestra el cuestionario diseñado para evaluar el riesgo de padecer una ECV en los próximos 10 años. Este cuestionario utiliza el modelo de clasificación desarrollado a partir del conjunto de datos de Framingham (National Heart, 2023), del cual obtuvimos una precisión del 85.9%. Los datos recopilados, abarcan información sobre el nivel educativo, adicciones, y antecedentes médicos. Al dar click en “Submit”

retorna como respuesta el tiene riesgo de contraer una ECV en los próximos 10 años o si no tiene riesgo de padecer una ECV en 10 años.

Los siguientes cuestionarios provienen de los modelos entrenados con el conjunto de datos combinado, datos del CDC de 2020 y 2022. Estos estos cuestionarios tienen la función de determinar si una persona está sufriendo de una ECV.

**Women Heart Stroke  
Consult Mixed Dataset**

Age:

Angina Type:

NAP

High Fasting Blood Sugar:

No

Resting ECG:

Normal

Resting ST depression:

Exercise Angina:

No

ST Slope Exercise:

Flat

Submit

**FIG 18.** Cuestionario de ECV de poblaciones variadas.

La figura 18 muestra el cuestionario derivado del modelo entrenado con el conjunto de datos combinados de cinco poblaciones distintas: Cleveland, Long Beach, Hungría, Suiza, y el conjunto ‘Statlog (Heart)’ del repositorio UC Irvine (Fedesoriano, 2021). Este modelo alcanzó una precisión del 100%, cuenta con 11 atributos comunes entre los distintos conjuntos de datos. El cuestionario está diseñado para detectar de manera precisa si un paciente está sufriendo actualmente de una enfermedad cardiovascular, utilizando la edad, la sintomatología, resultados de estudios de sangre, evaluación del

EKG, apoyando al equipo médico a brindar una evaluación del riesgo de estar padeciendo una ECV.

### Women Heart Failure 2020 Dataset

Smoking:

Stroke:

Difficulty Walking:

Age Category:

Diabetic:

General Health:

Asthma:

Kidney Disease:

**FIG 19.** Cuestionario con datos de ECV del CDC 2020.

La figura 19 muestra un cuestionario desarrollado haciendo uso del modelo entrenado, utilizando los datos del CDC para el año 2020 (Pytlak, 2020). Este modelo cuenta con una precisión del 70.51% y está específicamente diseñado para identificar si una paciente sufre de una ECV en el momento de la evaluación. El cuestionario incluye preguntas de hábitos de consumo de cigarrillos, si ha padecido un derrame cerebral, si tiene dificultad para caminar, la categoría de edad, si padece diabetes, como se encuentra su estado de salud, si padece asma y enfermedad de los riñones. Con esta información el modelo permite una detección temprana de ECV.

Age Category:  
18-24

Race Ethnicity:  
White

Height (Meters):

Smoking:  
Never smoked

Alcohol Consumption:  
No

Sleep Time (hours):

General Health:  
Poor

Last Check Up:  
Within past year

Removed Teeth:  
None

Had Angina:  
No

Had Stroke:  
No

Had Skin Cancer:  
No

Kidney Disease:  
No

Diabetic:  
No

HIV Test:  
No

Impaired Vision:  
No

Difficulty Walking:  
No

Flu Vaccination:  
No

Pneumo Vaccination:  
No

Submit

**FIG 20.** Cuestionario con datos de ECV del CDC 2022.

La figura 20 presenta el cuestionario desarrollado a partir del modelo entrenado con los datos del CDC 2022 (Pytlak, 2022), que alcanzó una precisión del 86%. Este cuestionario se enfoca en una evaluación detallada de la salud cardiovascular de la mujer. Este cuestionario considera la edad, cantidad de dientes removidos, salud en general, tiempo de sueño, raza, último chequeo médico, angina de pecho, estatura, padecer VIH, derrame cerebral, padecer cáncer de piel, el alcoholismo, las enfermedades de los riñones, la dificultad al caminar, el uso de cigarrillo, haber tenido vacunas de neumonía y de la gripe, y padecer diabetes para determinar si una mujer padece de alguna ECV.

**Women Heart Disease**  
**Lab Dataset**

Anemia:

No

Diabetes:

No

High Blood Pressure:

No

Serum Creatinine:

Serum Sodium:

Smoking:

No

Submit

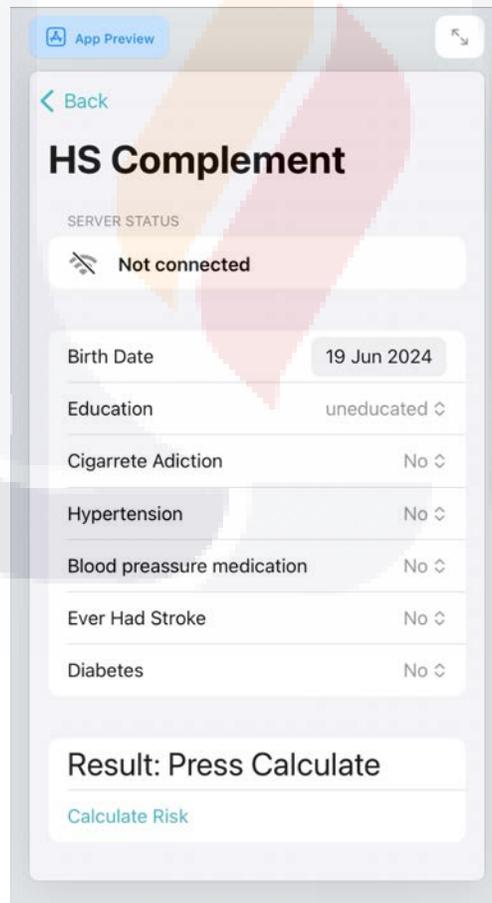
**FIG 21.** Cuestionario de mortalidad de IAM.

La figura 21 muestra un cuestionario desarrollado a partir del conjunto de datos de Punjab, Pakistán. El modelo de este cuestionario alcanzó una precisión del 100% en la evaluación del riesgo de mortalidad en mujeres que han sufrido un infarto. Este cuestionario está específicamente diseñado para analizar los factores que contribuyen a la mortalidad post-infarto en la mujer.

## Aplicación iOS

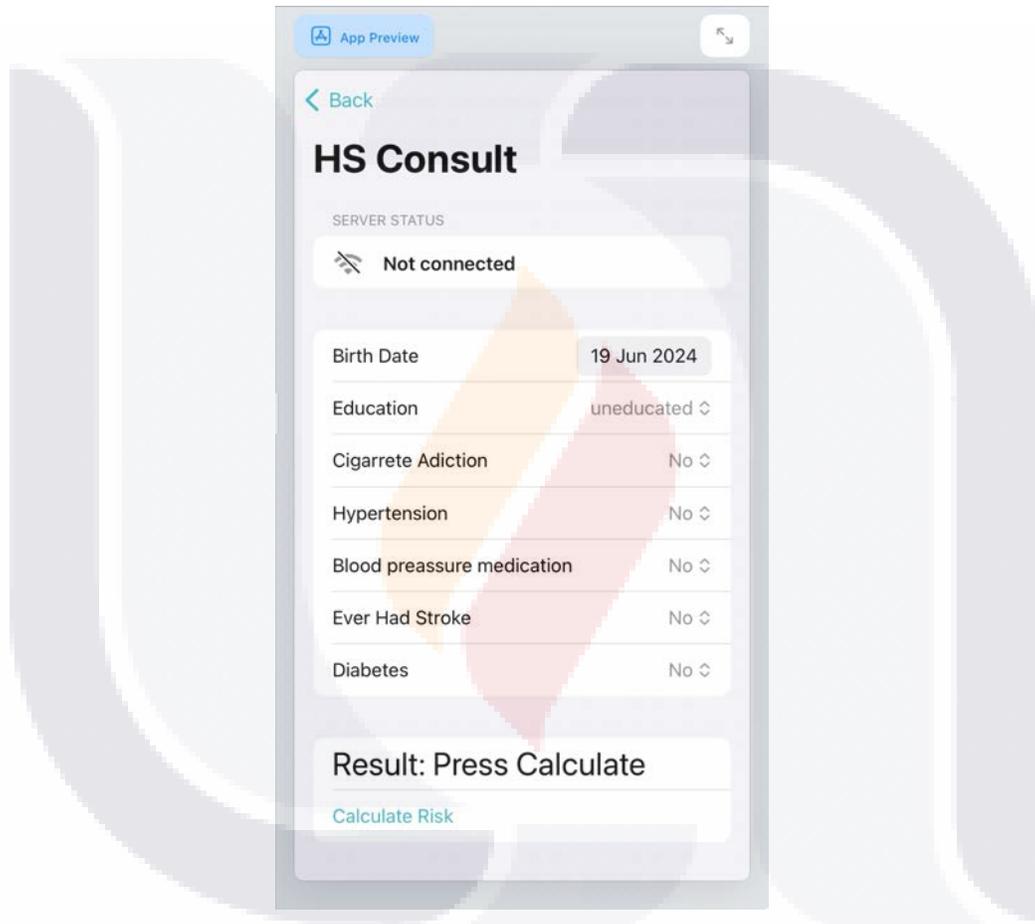
En esta sección, se presentan los cuestionarios desarrollados para la plataforma móvil iOS®, diseñados para evaluar diferentes aspectos de las ECV en mujeres. Estos cuestionarios se integran con el servidor web, lo que permite consultar y procesar los modelos predictivos a través de la red.

La aplicación, construida utilizando el marco de trabajo SwiftUI, maneja la comunicación con el servidor mostrando varios estados de conexión, tales como: desconectado, no se puede conectar al servidor, conectando, y conectado. Esta funcionalidad refleja en tiempo real el estado de la conexión a medida que la aplicación interactúa con el servidor en una red local. Este proyecto está desarrollado en el ambiente de desarrollo de Swift Playgrounds® y el proyecto se encuentra disponible en el siguiente repositorio de GitHub®: [https://github.com/rdls21/women\\_hd\\_ios](https://github.com/rdls21/women_hd_ios).



**FIG 22.** Cuestionario iOS: ECV en 10 años iOS.

Este cuestionario (Fig. 22) utiliza el modelo derivado del conjunto de datos de Framingham (National Heart, 2023) para estimar la probabilidad de que un paciente desarrolle una ECV en los próximos 10 años. Con una precisión del 85.9%, este cuestionario ofrece a los usuarios una herramienta valiosa para entender su riesgo futuro de enfermedad cardiovascular.



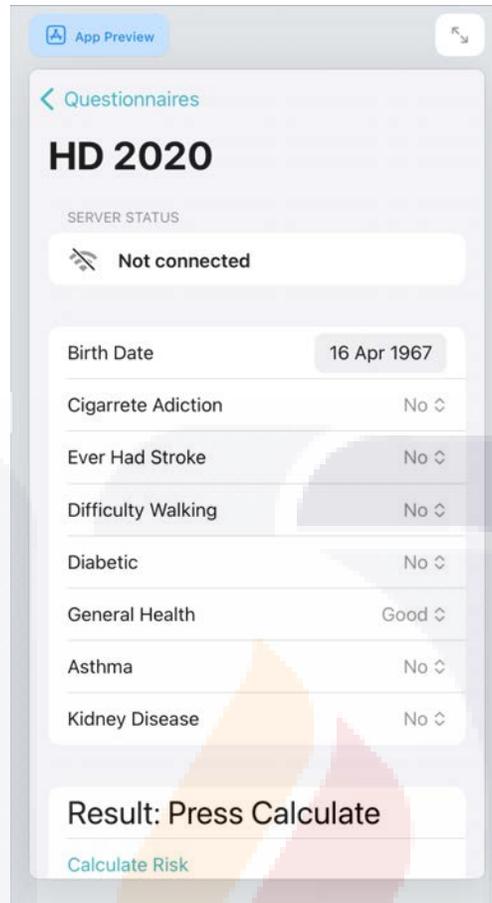
**FIG 23.** Cuestionario iOS: ECV de poblaciones variadas.

En la figura 23 se muestra el cuestionario desarrollado a partir del conjunto de datos combinado que incluye registros de Cleveland, Long Beach, Hungría, Suiza, y el conjunto "Statlog (Heart)" del repositorio UC Irvine (Fedesoriano, 2021). Cuenta con 12 características, incluido un indicador de ECV. Con una precisión del 100%, este modelo ayuda a identificar si una mujer esta sufriendo de una ECV.



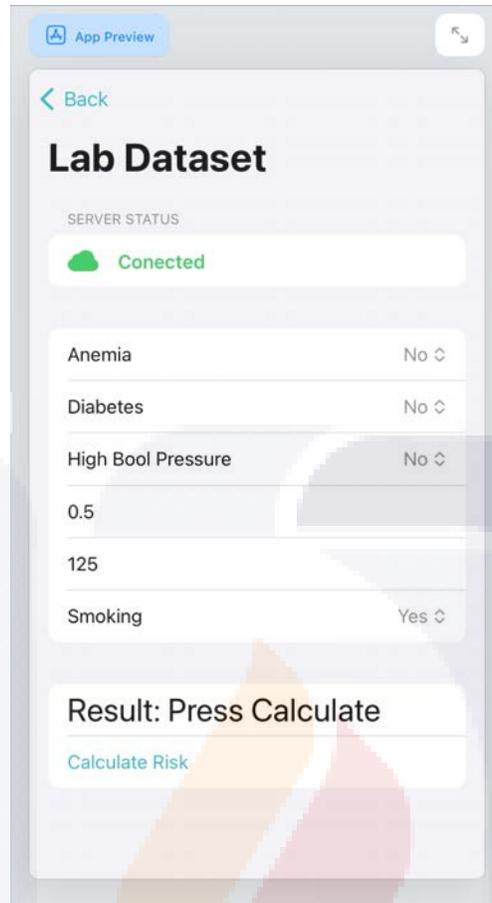
**FIG 24.** Cuestionario iOS: ECV del CDC 2020.

Utilizando datos del CDC de Estados Unidos para el año 2020, se desarrolló este cuestionario en la plataforma iOS (Figura 23), el cual tiene una precisión del 70.51%. Está diseñado para evaluar la presencia de ECV en mujeres. El cuestionario recopila información clave para determinar la probabilidad de que el usuario esté experimentando una ECV.



**FIG 25.** Cuestionario iOS: ECV del CDC 2022.

Con una precisión del 86%, este cuestionario se basa en los datos del CDC para el año 2022. Incluye una amplia gama de variables, tales como la edad, cantidad de dientes removidos, salud en general, tiempo de sueño, raza, último chequeo médico, entre otras. Este cuestionario es fundamental para identificar la presencia de ECV y evaluar el riesgo de los usuarios.



**FIG 26.** Cuestionario iOS: mortalidad de IAM.

Derivado de los datos del Instituto de Cardiología de Faisalabad en Punjab, Pakistán, este cuestionario evalúa la mortalidad en mujeres que han sufrido un infarto agudo de miocardio (IAM). Con una precisión del 100%, este modelo proporciona una predicción precisa del riesgo de mortalidad, ayudando a los profesionales de la salud a tomar decisiones informadas sobre la atención de las pacientes.

## IX. Discusión de Resultados

Los factores de riesgo obtenidos a partir de la implementación de esta metodología acentúan la importancia de incluir más factores de riesgo específicos para la mujer, ya que los pocos incluidos coinciden con los factores de riesgo sugeridos por la asociación canadiense del corazón (Mulvagh et al., 2022).

Para el conjunto de datos de Framingham (National Heart, 2023) obtuvimos un F1 de entrenamiento de 67%, lo cual sugiere que el modelo tiene un desempeño moderado, balanceando precisión y recall en los datos de entrenamiento. El modelo SVC tiene un rendimiento sólido en el conjunto de evaluación, con una alta exactitud del 84% y un Puntuación F1 de 85.96%, lo que refleja un equilibrio adecuado entre precisión y recall. El El Kappa de Cohen de 67.36% sugiere un buen nivel de acuerdo entre las predicciones del modelo y las etiquetas reales, ajustado por el azar. Además, un AUC de 86.98% siendo un modelo tiene con capacidad para distinguir entre las clases.

Para el conjunto de datos combinado (Fedesoriano, 2021) obtuvimos un Puntuación F1 en el entrenamiento de 80%, lo cual indica que el modelo tiene un buen desempeño en balancear precisión y recall. Los resultados obtenidos en el conjunto de evaluación son sobresalientes, con una exactitud del 100%, lo que significa que el modelo clasificó todas las instancias de manera correcta. La precisión y el recall también son del 100%, demostrando que el modelo no cometió errores en la predicción de las clases positivas y negativas. La Puntuación F1 del 100% muestra que el modelo alcanzó un equilibrio perfecto entre precisión y recall. Además, el El Kappa de Cohen de 100% indica que las predicciones del modelo y las etiquetas reales son idénticas. Por último, el AUC del 100% demuestra una capacidad de discriminación perfecta entre las dos clases.

Para el conjunto de datos del CDC 2020 (Pytlak, 2020), la regresión logística produjo un valor de F1 del 79% en el conjunto de entrenamiento, lo que sugiere que el modelo equilibra bien la precisión y el recall en esos datos. Este valor indica un desempeño robusto durante el entrenamiento, aunque no perfecto. La exactitud del modelo es del 75.75%, lo que refleja un buen rendimiento general, aunque por debajo del objetivo mínimo del 80%. La precisión del 70.51% revela la presencia de un número

significativo de falsos positivos, afectando así la precisión del modelo. Por otro lado, el recall del 85.49% indica que el modelo es bastante efectivo para identificar correctamente las instancias positivas.

Para el conjunto de datos del CDC 2022 (Pytlak, 2022). El modelo de Regresión Logística ha demostrado un rendimiento fuerte en el conjunto de evaluación, con una exactitud del 81.44%, lo que indica una alta proporción de predicciones correctas. La precisión del 86.01% y el Puntuación F1 del 81.13% reflejan un buen equilibrio entre precisión y recall, con una tendencia a minimizar los falsos positivos. Aunque el recall del 76.78% sugiere que el modelo identifica la mayoría de las instancias positivas, todavía hay espacio para mejorar en la detección completa. El El Kappa de Cohen del 63% refuerza la fiabilidad del modelo en sus predicciones, y un AUC de 89.4% refleja su capacidad para discriminar entre las clases. En resumen, el modelo de regresión logística proporciona un rendimiento sólido y fiable para la clasificación, con un buen equilibrio entre precisión y capacidad de identificación de instancias positivas.

Para el modelo entrenado con los datos de Punjab Pakistan (Chicco & Jurman, 2020), este modelo de regresión logística mostró un rendimiento excepcional en el conjunto de evaluación, alcanzando el 100% en todas las métricas evaluadas. Esto incluye la exactitud, precisión, recall, Puntuación F1, El Kappa de Cohen y el AUC, reflejando un buen desempeño, con capacidad para clasificar correctamente todas las instancias. Este modelo demostró una fiabilidad y precisión sobresalientes en la clasificación de los datos, garantizando una identificación precisa de las instancias evaluadas.

## **Recomendaciones**

Como aprendizaje de este trabajo se recomienda incluir factores de riesgo específicos en la mujer en estudios clínicos de todas las partes del mundo, debido a que la precisión en promedio, a pesar de que se lograron buenos resultados en los modelos, pudieron mejorar si se hubiesen tomado en cuenta en los conjuntos de datos que capturan las particularidades de la ECV en la mujer. También es importante explorar la posibilidad de utilizar estrategias de división de datos para el proceso de entrenamiento del modelo basado en bloques, lo asegura una búsqueda limitada de la mejor combinación de conjuntos de patos para el entrenamiento.

## X. Conclusiones

En conclusión, el uso de herramientas computacionales para la identificación de factores de riesgo en mujeres es factible haciendo uso de técnicas y herramientas computacionales.

Dado que en la mayoría de los modelos concebidos obtuvimos un resultado mayor al 80%. Consideramos que si es posible clasificar un infarto en las mujeres utilizando herramientas de aprendizaje automático.

También reconocemos la importancia de Support Vector Machine y Regresión Logística como algoritmos bien situados para clasificar correctamente las ECV en mujeres.

Haciendo uso de los factores de riesgo tradicionales si es posible clasificar el infarto en las mujeres, debido a que algunos conjuntos de datos no consideraban factores de riesgo emergentes y aún así lograron clasificar correctamente con una precisión mayor al 75%. Pero consideramos que el agregar factores de riesgo específicos en la mujer nos podrán brindar mejores resultados en el futuro.

El objetivo de este estudio en desarrollar una herramienta basada en aprendizaje automatizado, mediante la cual se identifique el riesgo de padecer ECV en mujeres se cumplió.

- Se generaron 5 modelos que permiten identificar los factores de riesgo específicos en la mujer.
- También se determinaron los dos algoritmos de aprendizaje automático para la problemática concreta.
- Se identificaron los factores de riesgo en la mujer a partir de los conjuntos de datos utilizados.
- Se desarrollo una aplicación WEB y móvil con las cuales se puede consultar si se está en riesgo de padecer una ECV en los próximos 10 años, si es alto el riesgo de padecer una ECV y la mortalidad de un infarto mediante un cuestionario.

## Trabajo Futuro

Una investigación posterior deberá profundizar aún más en la problemática que ocasiona esta enfermedad a las mujeres en Mexico, por lo que en futuros trabajos se continuaran utilizando técnicas inteligencia artificial, la estadística y el desarrollo de aplicaciones que permitan directamente a la mujeres tener una salud cardiovascular óptima, con el objetivo de disminuir la incidencia de esta enfermedad en la región. Para conseguir esto es indispensable considerar el crear una herramienta de recolección de datos, considerando los factores de riesgo específicos para la mujer.



## Bibliografía

5. Sex Affects Health. (2001). In T. M. Wizemann & M. L. Pardue (Eds.), *Exploring the Biological Contributions to Human Health: Does Sex Matter?* <https://doi.org/10.17226/10028>
- Aaronson, P. I., Ward, J. P., & Connolly, M. J. (2020). *The cardiovascular system at a glance*. John Wiley & Sons.
- Aggarwal, N. R., Patel, H. N., Mehta, L. S., Sanghani, R. M., Lundberg, G. P., Lewis, S. J., Mendelson, M. A., Wood, M. J., Volgman, A. S., & Mieres, J. H. (2018). Sex Differences in Ischemic Heart Disease. *Circulation: Cardiovascular Quality and Outcomes*, *11*(2), e004437. <https://doi.org/10.1161/CIRCOUTCOMES.117.004437>
- AHA, A. H. A. (2022). *Electrocardiogram (ECG or EKG)*. <https://www.heart.org/en/health-topics/heart-attack/diagnosing-a-heart-attack/electrocardiogram-ecg-or-ekg>
- Alty, S. R., Millasseau, S. C., Chowienzcyc, P. J., & Jakobsson, A. (2003, 27-30 Dec. 2003). Cardiovascular disease prediction using support vector machines. 2003 46th Midwest Symposium on Circuits and Systems,
- Álvarez Vega, M., Cortés Badilla, M. V., & Quirós Mora, L. M. (2021). El impacto de la microbiota en la enfermedad cardiovascular. *Revista Medica Sinergia*, *6*(2), e643. <https://doi.org/10.31434/rms.v6i2.643>
- Arnstein, P., Buselli, E. F., & Rankin, S. H. (1996). Women and heart attacks: prevention, diagnosis, and care. *The Nurse practitioner*, *21* 5, 57-58, 61-54, 67-59; quiz 70-51.
- Arroyo-Quiroz, C., Barrientos-Gutierrez, T., O'Flaherty, M., Guzman-Castillo, M., Palacio-Mejia, L., Osorio-Saldarriaga, E., & Rodriguez-Rodriguez, A. Y. (2020). Coronary heart disease mortality is decreasing in Argentina, and Colombia, but keeps increasing in Mexico: a time trend study. *BMC Public Health*, *20*(1), 162. <https://doi.org/10.1186/s12889-020-8297-5>
- Ashorobi, D., Ameer, M. A., & Fernandez, R. (2022). Thrombosis. In *StatPearls*. <https://www.ncbi.nlm.nih.gov/pubmed/30860701>
- Bader, F., Manla, Y., Atallah, B., & Starling, R. C. (2021). Heart failure and COVID-19. *Heart Fail Rev*, *26*(1), 1-10. <https://doi.org/10.1007/s10741-020-10008-2>
- Bhati, P., & Wunder, M. (2022). Women's Health: Atypical Presentation of Myocardial Infarction. *University of Western Ontario Medical Journal*.
- Breiman, L. (2001). Random forests. *Machine learning*, *45*, 5-32.
- Brown, J. C., Gerhardt, T. E., & Kwon, E. (2020). Risk factors for coronary artery disease.
- Chicco, D., & Jurman, G. (2020). Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone. *BMC Medical Informatics and Decision Making*, *20*(1), 16. <https://doi.org/10.1186/s12911-020-1023-5>
- Cho, L., Davis, M., Elgendy, I., Epps, K., Lindley, K. J., Mehta, P. K., Michos, E. D., Minissian, M., Pepine, C., Vaccarino, V., & Volgman, A. S. (2020). Summary of Updated Recommendations for Primary Prevention of Cardiovascular Disease in Women: JACC State-of-the-Art Review. *Journal of the American College of Cardiology*, *75*(20), 2602-2618. <https://doi.org/https://doi.org/10.1016/j.jacc.2020.03.060>
- Create ML | Apple Developer Documentation. (2023). Retrieved March 15 from <https://developer.apple.com/documentation/createml>
- Fedesoriano. (2021). *Heart Failure Prediction Dataset*. <https://www.kaggle.com/fedesoriano/heart-failure-prediction>
- Full, K. M., Huang, T., Shah, N. A., Allison, M. A., Michos, E. D., Duprez, D. A., Redline, S., & Lutsey, P. L. (2023). Sleep Irregularity and Subclinical Markers of Cardiovascular Disease: The Multi-Ethnic Study of Atherosclerosis. *J Am Heart Assoc*, *12*(4), e027361. <https://doi.org/10.1161/JAHA.122.027361>

- G, A., Ganesh, B., Ganesh, A., Srinivas, C., Dhanraj, & Mensinkal, K. (2022). Logistic regression technique for prediction of cardiovascular disease. *Global Transitions Proceedings*, 3(1), 127-130. <https://doi.org/https://doi.org/10.1016/j.gltp.2022.04.008>
- Grossman, W. (2024, 2024). *Heart Disease - Risk Factors*. University of California San Francisco. <https://healthyheart.ucsf.edu/risk-factors>
- Herrick, J. B. (1912). Landmark article (JAMA 1912). Clinical features of sudden obstruction of the coronary arteries. By James B. Herrick. *JAMA*, 250(13), 1757-1765. <https://www.ncbi.nlm.nih.gov/pubmed/6350634>
- Herrick, J. B. (1919). THROMBOSIS OF THE CORONARY ARTERIES. *Journal of the American Medical Association*, 72(6), 387-390. <https://doi.org/10.1001/jama.1919.02610060001001>
- Hunt, E. B. (2014). *Artificial intelligence*. Academic Press.
- INEGI, I. N. d. G. y. E. (2020). *Características De Las Defunciones Registradas En México Durante 2020, Preliminar* [https://www.inegi.org.mx/contenidos/saladeprensa/boletines/2021/EstSociodemo/DefuncionesRegistradas2020\\_Pre\\_07.pdf](https://www.inegi.org.mx/contenidos/saladeprensa/boletines/2021/EstSociodemo/DefuncionesRegistradas2020_Pre_07.pdf)
- INEGI, I. N. d. G. y. E. (2021). *México: principales causas de mortalidad en el estado de Aguascalientes 2021* [inegi.org.mx](https://inegi.org.mx)
- Ischemic Heart Disease. (2010). In C. o. S. S. C. D. Criteria (Ed.), *Cardiovascular Disability: Updating the Social Security Listings*. The National Academies. <https://doi.org/10.17226/12940>
- Jackins, V., Vimal, S., Kaliappan, M., & Lee, M. Y. (2021). AI-based smart prediction of clinical disease using random forest classifier and Naive Bayes. *The Journal of Supercomputing*, 77(5), 5198-5219. <https://doi.org/10.1007/s11227-020-03481-x>
- Johansson, B. W. (2001). [A history of the electrocardiogram]. *Dan Medicinhist Arbog*, 163-176. <https://www.ncbi.nlm.nih.gov/pubmed/11848076> (Elektrokardiografins utveckling.)
- Learning, E. (2023). *Clinical electrocardiography and ECG interpretation – ECG & ECHO*. <https://ecgwaves.com/topic/introduction-electrocardiography-ecg-book/>
- Leijdekkers, P., & Gay, V. (2008, 17-19 June 2008). A Self-Test to Detect a Heart Attack Using a Mobile Phone and Wearable Sensors. 2008 21st IEEE International Symposium on Computer-Based Medical Systems,
- Libby, P., & Theroux, P. (2005). Pathophysiology of coronary artery disease. *Circulation*, 111(25), 3481-3488. <https://doi.org/10.1161/CIRCULATIONAHA.105.537878>
- Lip, G. Y., Blann, A. D., Jones, A. F., & Beevers, D. G. (1997). Effects of hormone-replacement therapy on hemostatic factors, lipid factors, and endothelial function in women undergoing surgical menopause: implications for prevention of atherosclerosis. *Am Heart J*, 134(4), 764-771. [https://doi.org/10.1016/s0002-8703\(97\)70062-0](https://doi.org/10.1016/s0002-8703(97)70062-0)
- Lozano, R., Gómez-Dantés, H., Garrido-Latorre, F., Jiménez-Corona, A., Campuzano-Rincón, J. C., Franco-Marina, F., Medina-Mora, M. E., Borges, G., Naghavi, M., Wang, H., Vos, T., Lopez, A. D., & Murray, C. J. L. (2013). La carga de enfermedad, lesiones, factores de riesgo y desafíos para el sistema de salud en México [Article]. *Salud Pública de México*, 55(6), 580-594. <https://doi.org/10.21149/spm.v55i6.7304>
- Mehta, L. S., Beckie, T. M., DeVon, H. A., Grines, C. L., Krumholz, H. M., Johnson, M. N., Lindley, K. J., Vaccarino, V., Wang, T. Y., Watson, K. E., Wenger, N. K., American Heart Association Cardiovascular Disease in, W., Special Populations Committee of the Council on Clinical Cardiology, C. o. E., Prevention, C. o. C., Stroke, N., Council on Quality of, C., & Outcomes, R. (2016). Acute Myocardial Infarction in Women: A Scientific Statement From the American Heart Association. *Circulation*, 133(9), 916-947. <https://doi.org/10.1161/CIR.0000000000000351>

- Mirza, M. H. (2023). *Heart Disease Dataset - Framingham Heart Study* Kaggle. <https://doi.org/10.34740/KAGGLE/DSV/5146672>
- Mosca, L., Manson, J. E., Sutherland, S. E., Langer, R. D., Manolio, T., & Barrett-Connor, E. (1997). Cardiovascular disease in women: a statement for healthcare professionals from the American Heart Association. *Circulation*, *96*(7), 2468-2482.
- Mozaffarian, D., Benjamin, E. J., Go, A. S., Arnett, D. K., Blaha, M. J., Cushman, M., De Ferranti, S., Després, J.-P., Fullerton, H. J., & Howard, V. J. (2015). Heart disease and stroke statistics—2015 update: a report from the American Heart Association. *Circulation*, *131*(4), e29-e322.
- Mulvagh, S. L., Mullen, K.-A., Nerenberg, K. A., Kirkham, A. A., Green, C. R., Dhukai, A. R., Grewal, J., Hardy, M., Harvey, P. J., Ahmed, S. B., Hart, D., Levinsson, A. L. E., Parry, M., Foulds, H. J. A., Pacheco, C., Dumanski, S. M., Smith, G., & Norris, C. M. (2022). The Canadian Women's Heart Health Alliance Atlas on the Epidemiology, Diagnosis, and Management of Cardiovascular Disease in Women — Chapter 4: Sex- and Gender-Unique Disparities: CVD Across the Lifespan of a Woman. *CJC Open*, *4*(2), 115-132. <https://doi.org/https://doi.org/10.1016/j.cjco.2021.09.013>
- National Heart, L., and Blood Institute (NHLBI). (2023). *Framingham Heart Study*. Retrieved 19 de Marzo 2023 from <https://www.framinghamheartstudy.org/>
- Nielsen, J. D. (1995). Thrombogenesis. *Ann Chir Gynaecol*, *84*(4), 327-334. <https://www.ncbi.nlm.nih.gov/pubmed/8687075>
- NIH. (2022a). El sistema respiratorio - NHLBI [Health]. *NIH*. <https://www.nhlbi.nih.gov/es/salud/pulmones/sistema-respiratorio>
- NIH. (2022b). *What is atherosclerosis?* U.S. Department of Health and Human Services. <https://www.nhlbi.nih.gov/health/atherosclerosis>
- Pagan, E., Chatenoud, L., Rodriguez, T., Bosetti, C., Levi, F., Malvezzi, M., La Vecchia, C., Negri, E., & Bertuccio, P. (2017). Comparison of Trends in Mortality from Coronary Heart and Cerebrovascular Diseases in North and South America: 1980 to 2013. *Am J Cardiol*, *119*(6), 862-871. <https://doi.org/10.1016/j.amjcard.2016.11.040>
- Parmar, A., Katariya, R., & Patel, V. (2019, 2019//). A Review on Random Forest: An Ensemble Classifier. International Conference on Intelligent Data Communication Technologies and Internet of Things (ICICI) 2018, Cham.
- Prakash C Deedwania, M. (2022). Silent myocardial ischemia: Epidemiology, diagnosis, treatment, and prognosis. *UpToDate*. <https://www.uptodate.com/contents/silent-myocardial-ischemia-epidemiology-diagnosis-treatment-and-prognosis#>
- Purushottam, Saxena, K., & Sharma, R. (2015, 15-16 May 2015). Efficient heart disease prediction system using decision tree. International Conference on Computing, Communication & Automation,
- Pytlak, K. (2020). *Indicators of Heart Disease* Kaggle. <https://www.kaggle.com/datasets/kamilpytlak/personal-key-indicators-of-heart-disease>
- Pytlak, K. (2022). *Indicators of Heart Disease (2022 UPDATE)* Kaggle. <https://www.kaggle.com/datasets/kamilpytlak/personal-key-indicators-of-heart-disease>
- Rani, P., Kumar, R., Ahmed, N. M. S., & Jain, A. (2021). A decision support system for heart disease prediction based upon machine learning. *Journal of Reliable Intelligent Environments*, *7*(3), 263-275.
- Revueltas-Jiménez, F., Pleguezuelos-Ventura, A., Martínez-Montoya, F. J., Delgado-Barrios, C., & Redruello-Guerrero, P. (2022). Does ischemic cardiomyopathy behave differently in women? A holistic approach. *Actualidad Medica*, *106*(106(814)), 271-279. <https://doi.org/10.15568/am.2021.814.rev01>

- Rhcastilhos. (2010). Diagram of the human heart (cropped) pt. In D. o. t. h. h. c. pt.svg (Ed.). Wikimedia Commons: Wikipedia.
- Richard N. Fogoros, M. D. (2023). How the Heart's Electrical System Works. *Very Well Health*. <https://www.verywellhealth.com/cardiac-electrical-system-how-the-heart-beats-1746299>
- Rossello, X., Mas-Lladó, C., Pocock, S., Vicent, L., Van de Werf, F., Chin, C. T., Danchin, N., Lee, S. W. L., Medina, J., Huo, Y., & Bueno, H. (2021). Las diferencias por sexo en la mortalidad tras un síndrome coronario agudo se incrementan en los países de menor riqueza y mayor desigualdad de ingresos [10.1016/j.recesp.2021.05.009]. *Revista Española de Cardiología*. <https://doi.org/10.1016/j.recesp.2021.05.009>
- Salud, S. d. (2018). *Enfermedades cardiovasculares, gran reto de salud en México: Narro Robles* <https://www.gob.mx/salud/prensa/351-enfermedades-cardiovasculares-gran-reto-de-salud-en-mexico-narro-robles>
- Sambola, A., Elola, F. J., Ferreiro, J. L., Murga, N., Rodríguez-Padial, L., Fernández, C., Bueno, H., Bernal, J. L., Cequier, Á., Marín, F., & Anguita, M. (2021). Impacto de las diferencias de sexo y los sistemas de red en la mortalidad hospitalaria de pacientes con infarto agudo de miocardio con elevación del segmento ST. *Revista Española de Cardiología*, 74(11), 927-934. <https://doi.org/https://doi.org/10.1016/j.recesp.2020.07.031>
- Saxton, A., Tariq, M. A., & Bordoni, B. (2023). Anatomy, Thorax, Cardiac Muscle. In *StatPearls*. <https://www.ncbi.nlm.nih.gov/pubmed/30570976>
- Seber, G. A., & Lee, A. J. (2003). *Linear regression analysis* (Vol. 330). John Wiley & Sons. [https://books.google.com.mx/books?hl=es&lr=&id=mVSkEAAQBAJ&oi=fnd&pg=PR1&ots=UV6Nwtcyt\\_&sig=Ev1Q021Bvii09awnwNyCRXwKnk&redir\\_esc=y#v=onepage&q&f=false](https://books.google.com.mx/books?hl=es&lr=&id=mVSkEAAQBAJ&oi=fnd&pg=PR1&ots=UV6Nwtcyt_&sig=Ev1Q021Bvii09awnwNyCRXwKnk&redir_esc=y#v=onepage&q&f=false)
- Solorio, S., Hernández-González, M. A., Rangel Abundis, A., & Murillo-Ortiz, B. (2007). Cardiopatía isquémica en mujeres mexicanas. *Archivos de cardiología de México*, 77(3), 226-231.
- Srinivas, K., Rao, G. R., & Govardhan, A. (2010, 24-27 Aug. 2010). Analysis of coronary heart disease and prediction of heart attack in coal mining regions using data mining techniques. 2010 5th International Conference on Computer Science & Education,
- Suthaharan, S. (2016). Support vector machine. *Machine learning models and algorithms for big data classification: thinking with examples for effective learning*, 207-235.
- Tsao, C. W., Aday, A. W., Almarzooq, Z. I., Alonso, A., Beaton, A. Z., Bittencourt, M. S., Boehme, A. K., Buxton, A. E., Carson, A. P., Commodore-Mensah, Y., Elkind, M. S. V., Evenson, K. R., Eze-Nliam, C., Ferguson, J. F., Generoso, G., Ho, J. E., Kalani, R., Khan, S. S., Kissela, B. M., . . . null, n. (2022). Heart Disease and Stroke Statistics—2022 Update: A Report From the American Heart Association. *Circulation*, 145(8), e153-e639. <https://doi.org/10.1161/CIR.0000000000001052>
- Vancheri, F., Longo, G., Vancheri, E., & Henein, M. Y. (2022). Mental Stress and Cardiovascular Health—Part I. *Journal of Clinical Medicine*, 11(12), 3353. <https://doi.org/10.3390/jcm11123353>
- Wenger, N. K. (2012). Women and coronary heart disease: a century after Herrick: understudied, underdiagnosed, and undertreated. *Circulation*, 126(5), 604-611. <https://doi.org/10.1161/CIRCULATIONAHA.111.086892>
- WHO. (2020a). *The top 10 causes of death*. World Health Organization. Retrieved November 18 from <https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death>
- WHO. (2020b). *WHO reveals leading causes of death and disability worldwide: 2000-2019*. World Health Organization. Retrieved November 18 from <https://www.who.int/news/item/09-12-2020-who-reveals-leading-causes-of-death-and-disability-worldwide-2000-2019>

- WHO. (2022). *Deaths by sex and age group for a selected country or area and year* <https://platform.who.int/mortality/themes/theme-details/topics/topic-details/MDB/cardiovascular-diseases>
- Wu, J., Gale, C. P., Hall, M., Dondo, T. B., Metcalfe, E., Oliver, G., Batin, P. D., Hemingway, H., Timmis, A., & West, R. M. (2018). Editor's Choice - Impact of initial hospital diagnosis on mortality for acute myocardial infarction: A national cohort study. *European Heart Journal. Acute Cardiovascular Care*, 7(2), 139-148. <https://doi.org/10.1177/2048872616661693>
- Wu, O., Sumii, T., Asahi, M., Sasamata, M., Ostergaard, L., Rosen, B. R., Lo, E. H., & Dijkhuizen, R. M. (2007). Infarct prediction and treatment assessment with MRI-based algorithms in experimental stroke models. *J Cereb Blood Flow Metab*, 27(1), 196-204. <https://doi.org/10.1038/sj.jcbfm.9600328>
- Zhang, J., Jin, Y., Jia, P., Li, N., & Zheng, Z. J. (2021). Global Gender Disparities in Premature Death from Cardiovascular Disease, and Their Associations with Country Capacity for Noncommunicable Disease Prevention and Control. *Int J Environ Res Public Health*, 18(19). <https://doi.org/10.3390/ijerph181910389>
- Zhou, Z.-H. (2021). *Machine learning*. Springer Nature.





## A. Certificados y Constancias





UNIVERSIDAD  
CENTRO  
PANAMERICANO  
DE ESTUDIOS  
SUPERIORES

EL CONSEJO MEXICANO DE ESTUDIOS DE POSGRADO, A.C.  
Y LA UNIVERSIDAD CENTRO PANAMERICANO DE ESTUDIOS SUPERIORES

OTORGAN LA PRESENTE

# CONSTANCIA

A:

**Ricardo Daniel Lozano Sanchez**

Por haber asistido a la conferencia "La educación emocional en los ambientes de aprendizaje del siglo XXI", impartida por el Dr. Jorge Florentino Briceño González, de UNICEPES, en el marco del Programa Interinstitucional de Apoyo al Desarrollo Integral de Estudiantes de Posgrado 2023.2

01 septiembre del 2023

  
DR. ERICK SÁNCHEZ FLORES  
PRESIDENTE DEL COMEPO

  
M.C. JAZMÍN FLORES CHÁVEZ  
REPRESENTANTE INSTITUCIONAL











The Mexican Society for Artificial Intelligence (SMIA),  
 the Instituto de Investigaciones en Matemáticas Aplicadas y en  
 Sistemas, UNAM (IAE) and the Universidad Autónoma de  
 Yucatán (UADY)

Award this certificate to:

Lozano Sanchez, Ricardo Daniel; Torres Soto , María Dolores; Torres, Aurora; Sánchez Alemán ,  
 Esperanza

for presentation of the paper

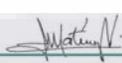
**Mortality Prediction on Adult and Old Women who Suffered a Myocardial Infarction**

at the workshop

**XVI workshop of Hybrid Intelligent Systems**

*held in conjunction with the 22<sup>nd</sup> Mexican International Conference on Artificial Intelligence (MICAI 2023)  
 Mérida, Yucatán, México, November 13 - 18, 2023.*

  
 Dr. Hiram Calvo  
SMIA President

  
 Dra. Lourdes Martínez  
Program Chair

  
 Dr. Hiram Ponce  
Program Chair

  
 Dr. Antonio Neme  
MICAI Local Chair








**La Sociedad Mexicana de Inteligencia Artificial (SMIA) y el Tecnológico Nacional de México - Instituto Tecnológico de Celaya**

**OTORGAN ESTE CERTIFICADO A**

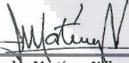
*Ricardo Daniel Lozano Sanchez, María Dolores Torres Soto, Aurora Torres, Yoselin Esparza Monreal, Cinthya Judith López Ramírez*

**por la presentación del artículo titulado**

*Myocardial Infarction Classification in Women*

**en el XVI Congreso Mexicano de Inteligencia Artificial - COMIA 2024  
Celaya, Guanajuato, México, del 3 al 7 de junio de 2024**



 Dra. Lourdes Martínez Villaseñor Presidenta SMIA	 Dra. Iris I. Méndez Gurrola Presidenta Comité de Programa	 Dr. Roberto A. Vázquez Espinosa Presidente Comité de Programa	 Dr. Alejandro I. Barranco Gutiérrez Comité Local COMIA
--	---	---	--

## B. Artículo

### Myocardial Infarction Classification in Women

Lozano Sánchez Ricardo Daniel<sup>1</sup>[10009-0007-4628-7827],  
Torres Soto María Dolores<sup>1</sup>[10000-0002-7245-1076], Torres Soto Aurora<sup>1</sup>[10000-0002-2930-824X],  
Yoselin Esparza Monreal<sup>2</sup>[10009-0009-6365-972X],  
Cinthya Judith López Ramírez<sup>2</sup>[10009-0002-2153-2180] and  
Sánchez Alemán Esperanza<sup>3</sup>[10000-0003-1745-858X]

<sup>1</sup> Dpto. de Sistemas de Información, Centro de Ciencias Básicas, Benemérita Universidad Autónoma de Aguascalientes. Av. Universidad, 940. CP 20131. Aguascalientes, Ags. México.

<sup>2</sup> Cardiología Clínica, Centenario Hospital Miguel Hidalgo. Avenida Gómez Morán, S/N Estación/Alameda. CP. 20259. Aguascalientes, Ags. México.

<sup>3</sup> Dpto. de Morfología, Centro de Ciencias Básicas, Benemérita Universidad Autónoma de Aguascalientes. Av. Universidad, 940. CP 20131. Aguascalientes, Ags. México.

a1236294@edu.uaa.mx,  
mdtorres@correo.uaa.mx, atorres@correo.uaa.mx,  
yosmonreal992@hotmail.com, dracinyaramirezchmh@gmail.com  
espesanchez@correo.uaa.mx

**Abstract.** Myocardial infarction is the leading cause of deaths in the world and women are less likely to survive from a myocardial infarction, we propose the use of two machine learning algorithms to train a classification model using clinical records of infarcted women to classify high mortality risk women and improve their life expectancy. We found women's clinical data can allow us to create a machine learning based model to predict mortality in 90% of infarcted patients. Using Apple's Create ML library we trained a Random Forest and Support Vector Machine based models using 105 records of hospitalized women from the Faisalabad Institute of Cardiology and the Allied Hospital in Faisalabad (Punjab, Pakistan). Our results shows that Random Forest outperforms Support Vector Machine with a 90% accuracy compared with the 85% accuracy of the Support Vector Machine model. This tool shows the capability of this kind of algorithms to create models that allow physicians to provide needed care to those infarcted women that are at risk of death.

**Keywords:** Myocardial infarction in women, Machine learning, Create ML.

#### 1. Introduction

According to the World Health Organization (WHO), myocardial infarction (MI) is one of the leading causes of deaths around the world, affecting nine million people in 2019, representing the 16% of all deaths globally (WHO, 2020). MI affects nearly 2.7 million women in the United States, this due to an increased prevalence rate of MI among older women and women of minority populations (Louisiana, 2016).

Women tend to experience atypical symptoms of MI and are often misdiagnosed or diagnosed too late, due to physiologic differences in a woman's heart that may pose a challenge to clinicians (Louisiana, 2016).

In Mexico, during the COVID-19 pandemic, the main cause of death in women was from heart-related diseases, exceeding 97,000 deaths from cardiovascular disease (CVD) compared to 70,000 deaths caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (INEGI, 2020), the following year, during January to July 2021, the main cause of deaths in women was also from CVD (INEGI, 2023).

Historically MI in women has been under-treated and understudied, leading to misdiagnosis and inadequate treatment (Wenger, 2012), also, women are less likely than men to receive treatment due to delays in the recognition of Acute Myocardial Infarction (AMI) by health providers or patients (Stehli et al., 2019). A study found that 18% of patients with non ST elevation myocardial infarction (NSTEMI) were given another non-heart related initial diagnosis, and women with ST elevation myocardial infarction (STEMI) had a 59% greater chance of an initial misdiagnosis compared with men (Wu et al., 2018) - (Kwok et al., 2021). A poor and delayed attention puts women at a high risk of mortality (Arnstein et al., 1996). These conditions are related to suboptimal awareness, sociocultural and financial reasons which may result in pre-hospital delays and lower rates of access to care. (Chandrasekhar et al., 2018).

The selected dataset contains 105 records of women who have experienced MI, these records span from 40 to 95 years old. It was published by Davide Chicco, Giuseppe Jurman, it contains data of 299 patients from the Faisalabad Institute of Cardiology and the Allied Hospital in Faisalabad (Punjab, Pakistan) during April to December of 2015. Their results showed that serum creatinine and ejection fraction can determine the mortality of a fractured patient, finding the random forest model (Breiman, 2001) performed with a 74% accuracy in their model validation (Chicco & Jurman, 2020). Although the authors of the dataset proposed serum creatinine and ejection fraction alone can predict the survival of the patients (Chicco & Jurman, 2020), we used more features contained in the dataset, this due to a weaker model validation score using only those features in previous experiments. For the creation of the classifier model we used a Random Forest (Breiman, 2001) and a Support Vector Machine (Amari & Wu, 1999) algorithms using Apple's Create ML tools (Create ML | Apple Developer Documentation, 2023).

The objective of this article is to develop a robust mortality risk classification tool focused on women who have experienced a MI using real clinical data. Our goal is to achieve an accuracy rate surpassing 80%. We seek to contribute with a valuable tool to the field of healthcare, reduce hospital mortality rates and improve life expectancy of patients with myocardial infarction. We delve into the development and viability of a classifier designed to distinguish women who have suffered a MI and assess their risk of mortality. This document contains a "theoretical framework" section, here we will discuss some heart related diseases, the MI disease, some risk and critical factors in women. After that, in the section "Material and method" we will discuss in depth the chosen dataset, the algorithms used and how the experiments will be taking place, in the "Results" section we will discuss any findings of the experimentation phase.

**2. Theoretical Framework**

Historically, heart-related diseases have been often perceived as primarily affection men, however, a significant body of research has found the impact of CVDs, such as coronary artery disease, heart failure, and stroke are the leading cause of mortality in women globally (Arroyo-Quiroz et al., 2020) (Fedorova et al., 2003) (WHO, 2022).

**2.1. Myocardial Infarction.**

A myocardial infarction is a condition in which the heart tissue suffers a lack of blood supply (ischemia), causing its death (necrosis); when a part of the heart tissue infarcts, it can produce severe complications to someone's health, including the risk of death. Common symptoms include chest pain, difficulty breathing, left arm or neck pain, among others. The cause of a myocardial infarction could be from several factors, such as: plaque accumulation, a cloth, or the vasoconstriction of the coronary arteries of the heart, resulting in the muscle to stop functioning, causing the death of the tissue ("Ischemic Heart Disease," 2010; Libby & Theroux, 2005).

**1. Gender Disparities in CVD.** Over the years, studies have revealed a deep gender disparities in the manifestation, diagnosis, and treatment of MI in women, they often exhibit unique symptoms that differ from the classic chest pain experienced by man, and, instead they might present with subtle signs as shortness of breath, nausea, fatigue or confused with gastrointestinal, anxiety or emotional symptoms ("5. Sex Affects Health," 2001; Chandrasekhar et al., 2018). This divergence of symptoms has led to under-treatment, under-diagnosis and delayed treatment (Vest & Cho, 2020; Wenger, 2012).

**2. Risk Factors and Underlying Causes of MI.** Heart-related diseases in women have additional risk factors to those that are common in men, such as complications during pregnancy, hormonal fluctuations and chronic stress ("5. Sex Affects Health," 2001). Understanding those distinct risk factors is essential for accurate risk assessment and allows to develop prevention strategies specifically for woman (Chandrasekhar et al., 2018).

**3. Hormonal Changes.** Hormones such as estrogens contribute significantly to women's health, making it subject of extensive research concerning their heart health. Estrogen offers protective cardiovascular effects, such as maintaining healthy levels of cholesterol in blood, women who have lower levels of estrogen, specifically during menopause, are more prone to CVD, including AMI (Fedorova et al., 2003; Louisiana, 2016).

**2.2. Artificial Intelligence**

AI is the simulation of human intelligence in machines, capable of learning, problem-solving and thinking like humans. AI encompasses a wide range of technologies and techniques applicable to real world problems and tasks, aiming specifically to tasks that typically require human intelligence, such as tasks requiring the understanding and capability of natural language, image recognition, solving complex problems, and making decisions (Hunt, 2014).

**1. Machine Learning.** ML is a technique used by artificial intelligence that focuses on the development of algorithms that allow computers to learn, make predictions and classifications based on data. ML algorithms use statistical techniques to learn patterns from large datasets. ML algorithms can be categorized into three types: Supervised Learning: the model is trained on labeled data, Unsupervised Learning: model identifies patterns and structures, such as clustering, and Reinforcement Learning: learns to make decisions by interacting with the environment, it receives feedback in the form of rewards or penalties (Zhou, 2021).

**2. Random Forest Algorithm.** RF is an ML supervised algorithm used for regression and classification tasks, belongs to the ensemble learning methods. It combines predictions from multiple decision trees algorithms to make a more accurate and robust prediction than any individual model. Also has a key feature to randomly select each feature for the nodes of each tree, de-correlates the trees and ensures a diverse and robust prediction (Breiman, 2001) it is currently used in healthcare, finance, ecology and remote sensing (image classification) (Al-Manaseer et al., 2022; Belgiu & Drăguț, 2016; Cutler et al., 2007).

**3. Support Vector Machine Algorithm.** SVM is a supervised ML algorithm used for classification and regression tasks. It is well suited for data where clear margins of separation can be drawn between different classes or groups, it finds the optimal hyperplane that best separates different classes with the maximum margin, which is the distance between the hyperplane and the nearest data points that separates each class. Maximizing the margin ensures a better generalization and robustness of the model. Also the kernel trick allows linear and non-linear data to be handled by transforming the input features into a higher dimensional space using kernel functions (Amari & Wu, 1999; Suthaharan & Suthaharan, 2016).

**4. Factorial Experiment Design.** Is an approach used in experimental research to study the influence of multiple variables simultaneously. This method allows to explore the individual effects of several variables on a dependent variable. As its core concepts we have factors: denoted as the variables being studied, Levels as the different values of those factors and combinations (Spall, 2010; Yates, 1937).

### 3. Material and Method

We present a detailed account of the methodologies employed in our study on the classification of adult women at risk of mortality after having a MI, using a machine learning technique and using a SVM and RF algorithms. The following overview of the execution of this experiment, tools, and information used in this are described in the following points.

#### 3.1. Dataset

This dataset published along the paper “Serum Creatinine And ejection Fraction can predict survival of patients from heart failure” contains 105 women and 194 men, ranging from 40 to 95 years old, this dataset was created using data collected at the Faisalabad Institute of Cardiology and at the Allied Hospital in Faisalabad (Punjab, Pakistan) during April to December of 2015. From all 299 patients only 96 died (32.10%), more specifically, from all 105 women only 34 died, which corresponds to a 32.38% of deaths in the women dataset, on the other hand, from the 194 man, 62 died which corresponds to a 31.95%. All 299 patients had left ventricular systolic dysfunction and had previous heart failures (Chicco & Jurman, 2020).

Table 1. Dataset Features

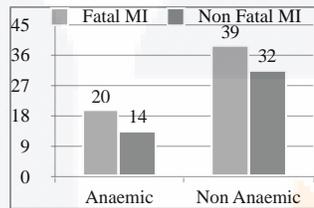
Feature Name	Description
Age	The age of the patient
Anaemia	Boolean value depending whether the patient has anemia
Creatinine Phosphokinase	Levels of CPK enzyme in blood, mcg/L
Diabetes	Boolean value depending whether the patient has diabetes
Ejection Fraction	Percentage of blood leaving the hearth at each contraction
High Blood Pressure	Boolean value whether the patient has high blood pressure
Platelets	Platelets count in blood, kilo-platelets/mL
Serum Creatinine	Level of serum creatinine in the blood mg/dl
Serum Sodium	Level of serum sodium in the blood mEq/L
Sex	Binary value whether the patient is a woman or man
Smoking	Boolean value whether the patient smokes
Time	The follow-up period in days
Death Event	Whether the patient deceased during the follow up period

Features such as Sex and Time (Table 1) are not necessary since we will focus only on women patients and those features do not provide valuable information.

6

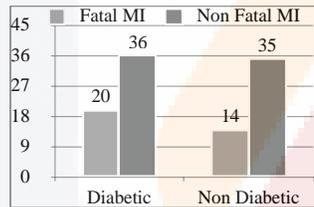
**3.2. Description of the dataset**

We compare a set of categorical features versus the dependent variable, this process is made to have a better understanding of the dataset.



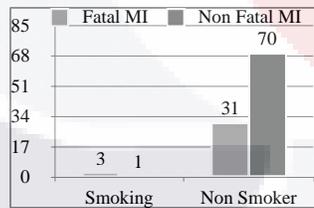
**Fig. 1** Death event

The Figure 1 makes visible that infarcted women are less prone to death whether they are diagnosed anaemic, thus the outcome could not be described by this characteristic.



**Fig. 2** Death event

As the Figure 2 shows, infarcted women are more prone to death whether they are diagnosed diabetic, though the outcome is either described by this characteristic.



**Fig. 3** Death event

This Figure 3, reveals that smoker women does not present a mayor indicator of a fatal MI in this population, although it is consider a major risk factor.

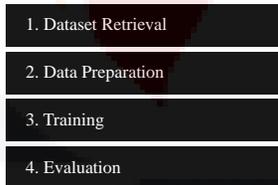
**3.3. Algorithms**

Apple’s Create ML is a set of machine learning algorithms designed specifically for developers using macOS and iPadOS platforms. Introduced by Apple in 2019; this framework allows developers to build, train and test machine learning models without requiring previous experience in machine learning algorithms. Along Xcode, it can be used either with an easy-to-use application or directly while coding using their Swift Library. (Create ML | Apple Developer Documentation, 2023).

**1. Algorithms Selection.** According to the article “Enhanced cardiovascular disease prediction model using random forest algorithm”. RF achieved an accuracy of more than 99% by classifying patients with CVD, outperforming SVM, Logistic Regression, K-Means, among other models (Al-Manaseer et al., 2022). Although, SVM did not perform as good as RF; the article “Early Coronary Heart Disease Deciphered via Support Vector Machines: Insights from Experiments” demonstrate that SVM is a suitable algorithm for a patient dataset with Coronary Heart Disease (CHD), being capable of classify CHD patients with a precision of 87.8% using the sigmoid kernel (Akhtar et al., 2023). We decided to use the Create ML implementation for a SVM and a RF algorithms. The SVM and RF algorithm within Create ML provides efficient learning models, complemented the ease of use and that it takes advantage of the capacity of Apple’s ARM architecture to efficiently train ML models (Jagadeesh, 2022).

**3.4. Experiments**

The aforementioned dataset is used for both the SVM and the RF algorithms to train a model, first we carry out an empirical study to find the best set of parameters for each algorithm, then, based on the results of the previous phase, an exhaustive parameter search is performed using a factorial experiment design, then the model is evaluated.



**Fig. 4.** Methodology

The figure 4 represents the steps taken to design, train and evaluate the model using both algorithms, SVM and RF.

**1. Dataset Retrieval.** The dataset was obtained during research aimed at predicting heart infarcts using a machine learning-based application, focusing on women. This dataset was sourced from the paper “Machine Learning Can Predict Survival of Patients with Heart Failure from Serum Creatinine and Ejection Fraction Alone” by Chicco and Jurman (2020). This data comprises clinical information collected from patients at the Faisalabad Institute of Cardiology in Faisalabad (Punjab, Pakistan) between April and December, 2015. Among the 299 records in this data set, 105 belong to women, of which 34 unfortunately died during the study period.

**2. Data Preparation.** Our focus is exclusively on female cases; to ensure this, we have filtered the full dataset to include only female records, creating a new dataset specifically adapted for the purpose of this article. To maintain an even distribution of the 'DEATH\_EVENT' feature, ensuring an approximate 3:10 ratio of deaths to survival, we carefully curated our dataset. This balanced dataset is it's divided into the training and testing splits, which serves as the foundation for our efforts during the following steps. We established that the training split will comprise 80% of this filtered dataset, and the testing split, encompasses the remaining 20%. In order to remove non correlative features, we further refined this dataset, we removed the 'sex' and 'time' features, as the sex did not provide any significant data, since all remaining records where 'woman' and follow-up period does not correlate with an individual's mortality in this context.

**3. Training.** Our initial approach involved a empirical exploration to identify optimal parameters for Support Vector Machine (Amari & Wu, 1999) and Random Forest (Breiman, 2001), after finding a promising set of parameters for each algorithm individually, we employed a factorial experiment design (Spall, 2010) using a wide range of parameters, each result of every combination was recorded on a CSV file, enabling us to find the most effective parameters configuration. During this phase, it was observed that Create ML automatically generates a validation split. Although Apple has not disclosed specific details, such as the threshold or percentage used in this split mechanism (Create ML | Apple Developer Documentation, 2023), we observed that it serves as feedback on every iteration during the training phase, improving the models accuracy.

**4. Evaluation.** In this phase, we asses the results obtained from the previous phrase, reviewing the created files with the accuracy for each combination result of the Factorial Experiment Design and involved rigorous testing and optimization for our machine learning models, with this insights.

In the following section we will present the results of this methodology.

#### 4. Results

First, we began with the SVM algorithm, scoring 85% of accuracy after evaluating the model using the testing dataset, after that the Random Forest algorithm gave the best evaluation score with 90% of accuracy on the testing dataset, in the following sections we discuss how those experiments were performed.

##### 4.1. Support Vector Machine

Testing the SVM max iterations parameter as the convergence parameter, we found that 100 maximum iterations gave us promising results, this values where enough for the algorithm to provide a mature model with promising results, although, testing with some values it needed more iterations; we decided to also test with 1,000 maximum iterations, for the penalty parameter we found that values from 25 to 65 the algorithm was giving promising results. The convergence threshold was kept to a low value of 0.001 to assure great results, the feature rescaling parameter was set to TRUE to all experiments.

Table 2. Support Vector Machine Parameters for the Factorial Experiment Design

Parameters	Values	k value
Max Iterations	100 and 1,000	$k_1 = 2$
Penalty	25 to 65, steps by 0.1	$k_2 = 400$
Convergence Threshold	0.001	$k_3 = 1$
Feature Rescaling	TRUE	$k_4 = 1$

The features used in the factorial experimentation design are displayed on table 2.

$$30(2 * 400) = 24,000 \tag{1}$$

As the formula (1) displays, the number of iterations of the factorial experiment are 24,000. This experiment gave us the best result as disclosed on the Table 3.

Table 3. Factorial Experiment Design Results for Support Vector Machine

Parameters	Confusion Matrix	Precision									
Max Iterations: 100 Penalty: 35.0 Convergence Threshold: 0.001 Feature Rescaling: true	<table border="1"> <tr> <td>True\Pred</td> <td>false</td> <td>true</td> </tr> <tr> <td>false</td> <td>14</td> <td>0</td> </tr> <tr> <td>true</td> <td>3</td> <td>4</td> </tr> </table>	True\Pred	false	true	false	14	0	true	3	4	85.71%
True\Pred	false	true									
false	14	0									
true	3	4									

Results in Table 3 shows a well-matured model, but do not meet our expectations.

**4.2. Random Forest**

When random forest algorithm was trained with a thousand maximum iterations to converge gave us worse results since it was the stop parameter. We found that the number of iterations that gave us promising results for this parameter was between 19 and 120 maximum iterations, For the maximum depth parameter, the best results we obtained from 4 and the algorithm stopped to give good results to the 6 depth level. The min-loss reduction and the min-child weight parameters gave the best results between 0.001 and 0.4. Lastly, the row subsample and column subsample parameters performed the best at 0.8.

We conducted extensive iterations to find the optimal parameters for convergence. Surprisingly, the use of a thousand iterations led to bad results. We discovered that the sweet spot for the number of iterations, providing the most promising outcomes, ranged between 19 and 120. Similarly, in determining the maximum depth parameter, we found that the algorithm consistently produced favorable results within the range of 4 to 6 levels. Specifically, the algorithm automatically ceased to yield significant improvements beyond the 6th level. Additionally, fine-tuning the 'min loss reduction' and 'min child weight' parameters proved critical. The best results were achieved when these values were set between 0.001 and 0.4. Moreover, we observed optimal performance when setting 'row subsample' and 'column subsample' parameters at 0.8. These findings highlight the importance of meticulous parameter tuning in harnessing the true potential of the random forest algorithm.

Table 4. Random Forest Parameters for the Factorial Experiment Design

Parameters	Values	<i>k</i> value
Max Depth	4 to 6, steps of 1	$k_1 = 3$
Max Iterations	15 to 130, steps of 15	$k_2 = 8$
Min Loss Reduction	0.001 to 0.41, steps by 0.05	$k_3 = 9$
Min Child Weight	0.001 to 0.41, steps by 0.05	$k_4 = 9$
Row Subsample	80%	$k_5 = 1$
Column Subsample	80%	$k_6 = 1$
Random Seed	46	$k_7 = 1$

Taking those results on account, the parameters used in the factorial experimentation design (Spall, 2010; Yates, 1937) are displayed on the Table 4.

$$30(3 * 8 * 9 * 9) = 58,320 \tag{2}$$

As the formula (2) displays, the number of iterations of the factorial experiment which are 58,320. In our current setting it took several hours to complete, but finally the results:

Table 5. Factorial Experiment Design Results for Random Forest

Parameters	Confusion Matrix			Precision
Max Depth: 4				
Max Iterations: 60				
Min Loss Reduction: 0.001	True\Pred	false	true	
Min Child Weight: 0.001	false	14	0	90.48%
Random Seed: 46	true	2	5	
Row Subsample: 0.8				
Column Subsample: 0.8				

This final results showed in Table 5 that RF outperformed SVM with better precision of the model compared with the results from SVM Table 3 but as it is in the case of the SVM, as seen in the confusion matrix from both model results (Table 3 and Table 5) the models struggle at identifying true cases, this due to the poor exposition of the algorithms to true cases, making it weaker in this class.

In this case the RF algorithm outperformed SVM, this could be due to the narrow search space for SVM, and coincidentally the search space for the hyper-parameters of RF gave good results without the need to change the search space.

**5. Conclusions**

This model works effectively within the defined scope, making precise predictions mainly for patients within similar age groups. Additionally, it's crucial to recognize that race factors play a significant role in influencing the model's accuracy and a death ratio of around 3:10 showed as a better performance classifying false class on both model's confusion matrix (Table 3 and Table 5). Acknowledging the dataset nuances is essential, since they demonstrate the need to understand the specific characteristics that vary depending on the demographics of each country..

**5.1. Future work**

The expansion of the scope of this medical classifier entails the potential to improve the life expectancy of infarcted women. Accounting data from a larger and a diverse pool of patients, it is possible to capture a broader spectrum of health profiles, this could enable the model to recognize intricate patterns and correlations. Widening the age range will allows to account for the unique physiological factors in different life stages. Additionally, considering a wider geographical region introduces variables that could significantly impact health outcomes.

**Bibliography**

5. Sex Affects Health. (2001). In T. M. Wizemann & M. L. Pardue (Eds.), *Exploring the Biological Contributions to Human Health: Does Sex Matter?* <https://doi.org/10.17226/10028>

Akhtar, F., Heyat, M. B. B., Parveen, S., Singh, P., Hassan, M. F. U., Parveen, S., Hayat, M. A. B., Sayeed, E., Ali, A., Li, J. P., & Sawan, M. (2023, 15-17 Dec. 2023). Early Coronary Heart Disease Deciphered via Support Vector Machines: Insights from Experiments. 2023 20th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP).

Al-Manaseer, H., Abualigah, L., Alsoud, A. R., Zitar, R. A., Ezugwu, A. E., & Jia, H. (2022). A novel big data classification technique for healthcare application using support vector machine, random forest and J48. In *Classification applications with deep learning and machine learning technologies* (pp. 205-215). Springer.

Amari, S.-i., & Wu, S. (1999). Improving support vector machine classifiers by modifying kernel functions. *Neural Networks, 12*(6), 783-789. <https://www.sciencedirect.com/science/article/abs/pii/S0893608099000325>

Arnstein, P., Buselli, E. F., & Rankin, S. H. (1996). Women and heart attacks: prevention, diagnosis, and care. *The Nurse practitioner, 21* 5, 57-58, 61-54, 67-59; quiz 70-51.

Arroyo-Quiroz, C., Barrientos-Gutierrez, T., O'Flaherty, M., Guzman-Castillo, M., Palacio-Mejia, L., Osorio-Saldarriaga, E., & Rodriguez-Rodriguez, A. Y. (2020). Coronary heart disease mortality is decreasing in Argentina, and Colombia, but keeps increasing in Mexico: a time trend study. *BMC Public Health, 20*(1), 162. <https://doi.org/10.1186/s12889-020-8297-5>

Belgiu, M., & Drăguț, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS journal of photogrammetry and remote sensing, 114*, 24-31.

Breiman, L. (2001). Random forests. *Machine learning, 45*, 5-32.

Chandrasekhar, J., Gill, A., & Mehran, R. (2018). Acute myocardial infarction in young women: current perspectives. *Int J Womens Health, 10*, 267-284. <https://doi.org/10.2147/IJWH.S107371>

Chicco, D., & Jurman, G. (2020). Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone. *BMC Medical Informatics and Decision Making, 20*(1), 16. <https://doi.org/10.1186/s12911-020-1023-5>

Create ML | Apple Developer Documentation. (2023). Retrieved March 15 from <https://developer.apple.com/documentation/createml>

Cutler, D. R., Edwards Jr, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). Random forests for classification in ecology. *Ecology, 88*(11), 2783-2792.

Fedorova, E. L., Bondareva, Z. G., Kuimov, A. D., & Nesterenko, E. V. (2003). Myocardial infarction in women: risk factors and clinical features. *Klin Med (Mosk), 81*(6), 28-32. (Faktoy riska i osobennosti techeniia infarkta miokarda u zhenshchin.)

Hunt, E. B. (2014). *Artificial intelligence*. Academic Press.

INEGI, I. N. d. G. y. E. (2020). *Características De Las Defunciones Registradas En México Durante 2020, Preliminar* [https://www.inegi.org.mx/contenidos/saladeprensa/boletines/2021/EstSociodemo/DefuncionesRegistradas2020\\_Pre\\_07.pdf](https://www.inegi.org.mx/contenidos/saladeprensa/boletines/2021/EstSociodemo/DefuncionesRegistradas2020_Pre_07.pdf)

INEGI, I. N. d. G. y. E. (2023). *Estadística de Defunciones Registradas de Enero a Junio de 2022 (Preliminar)* <https://www.inegi.org.mx/contenidos/saladeprensa/boletines/2023/DR/DR-EnE-jun2022.pdf>

Ischemic Heart Disease. (2010). In C. o. S. S. C. D. Criteria (Ed.), *Cardiovascular Disability: Updating the Social Security Listings*. The National Academies. <https://doi.org/10.17226/12940>

Kwok, C. S., Bennett, S., Azam, Z., Welsh, V., Potluri, R., Loke, Y. K., & Mallen, C. D. (2021). Misdiagnosis of acute myocardial infarction: A systematic review of the literature. *Critical Pathways in Cardiology*, 20(3), 155-162.

Libby, P., & Theroux, P. (2005). Pathophysiology of coronary artery disease. *Circulation*, 111(25), 3481-3488. <https://doi.org/10.1161/CIRCULATIONAHA.105.537878>

Louisiana, L. W. P. M. C. A. P. E. M. P. C. C. G. P. B. C. A. P. X. U. o. L. C. o. P. N. O. (2016). Myocardial Infarction and Older Women. <https://www.uspharmacist.com/article/myocardial-infarction-and-older-women>

Spall, J. C. (2010). Factorial Design for Efficient Experimentation. *IEEE Control Systems Magazine*, 30(5), 38-53. <https://doi.org/10.1109/MCS.2010.937677>

Stehli, J., Martin, C., Brennan, A., Dinh, D. T., Lefkovits, J., & Zaman, S. (2019). Sex Differences Persist in Time to Presentation, Revascularization, and Mortality in Myocardial Infarction Treated With Percutaneous Coronary Intervention. *J Am Heart Assoc*, 8(10), e012161. <https://doi.org/10.1161/JAHA.119.012161>

Suthaharan, S., & Suthaharan, S. (2016). Support vector machine. *Machine learning models and algorithms for big data classification: thinking with examples for effective learning*, 207-235.

Vest, A. R., & Cho, L. (2020). No Woman Left Behind. *Circulation: Heart Failure*, 13(10), e007782. <https://doi.org/doi:10.1161/CIRCHEARTFAILURE.120.007782>

Wenger, N. K. (2012). Women and coronary heart disease: a century after Herrick: understudied, underdiagnosed, and undertreated. *Circulation*, 126(5), 604-611. <https://doi.org/10.1161/CIRCULATIONAHA.111.086892>

WHO. (2020). *The top 10 causes of death*. World Health Organization. Retrieved November 18 from <https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death>

WHO. (2022). *Deaths by sex and age group for a selected country or area and year* <https://platform.who.int/mortality/themes/theme-details/topics/topic-details/MDB/cardiovascular-diseases>

Wu, J., Gale, C. P., Hall, M., Dondo, T. B., Metcalfe, E., Oliver, G., Batin, P. D., Hemingway, H., Timmis, A., & West, R. M. (2018). Editor's Choice - Impact of initial hospital diagnosis on mortality for acute myocardial infarction: A national cohort study. *Eur Heart J Acute Cardiovasc Care*, 7(2), 139-148. <https://doi.org/10.1177/2048872616661693>

Yates, F. (1937). The design and analysis of factorial experiments.

Zhou, Z.-H. (2021). *Machine learning*. Springer Nature.