

UNIVERSIDAD DE CIENCIAS MÉDICAS DE SANTIAGO DE CUBA

UNIVERSIDAD DE CIENCIAS MÉDICAS DE HOLGUÍN

POLICLÍNICO DOCENTE

“JOSÉ ÁVILA SERRANO”

**MODELO PREDICTIVO DE ENFERMEDAD CARDIOVASCULAR BASADO EN
INTELIGENCIA ARTIFICIAL EN LA ATENCIÓN PRIMARIA DE SALUD**

Tesis presentada en opción al grado científico de

Doctor en Ciencias Médicas

JORGE BAUDILIO VEGA ABASCAL

Holguín, 2025

UNIVERSIDAD DE CIENCIAS MÉDICAS DE SANTIAGO DE CUBA

UNIVERSIDAD DE CIENCIAS MÉDICAS DE HOLGUÍN

POLICLÍNICO DOCENTE

“JOSÉ ÁVILA SERRANO”

**MODELO PREDICTIVO DE ENFERMEDAD CARDIOVASCULAR BASADO EN
INTELIGENCIA ARTIFICIAL EN LA ATENCIÓN PRIMARIA DE SALUD**

Tesis presentada en opción al grado científico de
Doctor en Ciencias Médicas

Autor: Profesor Auxiliar, Dr. Jorge Baudilio Vega Abascal, Ms.C.

Tutores: Profesor Titular, Dr. Alberto Rubén Piriz Assa, Dr.C.

Profesor Titular, Dr. Diego Nápoles Riaño Dr.C.

Holguín, 2025

PENSAMIENTO

“La ciencia médica se desarrolla constantemente. Como resultado de la experiencia, de la investigación, surgen nuevos métodos, nuevas técnicas; nosotros tenemos que apoderarnos de esas técnicas”.

A handwritten signature in black ink, reading "Fidel Castro Ruz", enclosed within a simple, elongated oval outline.

Fidel Castro Ruz
12 de marzo de 1982



TESTIMONIOS DE GRATITUD

A Dios y a la patrona de Cuba, la Virgen María de la Caridad del Cobre, porque mucha ciencia nos acerca a Dios, poca, nos aleja, como dijo Pasteur. Dios nos muestra el camino cuando no podemos verlo, nos esclarece los objetivos, porque cuando hayamos hecho todo lo que está a nuestro alcance, entonces Dios hará el resto.

Al Dr.C. Rafael Trinchet Soler, por su sabiduría y dedicación, motor impulsor del programa doctoral en Holguín y a doctores y doctorandos de ese gran equipo por ofrecer siempre su inestimable ayuda.

A mis tutores Dr.C. Piriz Assa y Dr.C. Diego Nápoles Riaño, ya un gran equipo, por haber confiado en mí, su inmenso apoyo y guía en estos años de duro bregar imponiéndonos a dificultades, sorteando obstáculos, pero siempre saliendo adelante.

A los profesores Dr. Salvador Tamayo Muñiz, Coordinador Nacional de ENT del MINSAP; Dr.C. Jorge Luis León Álvarez, presidente de la Comisión Nacional Técnica Asesora de HTA; y al Dr. Baster Moro, colega de mil batallas, por su interés en la investigación. Al Dr.C. Rolando Bonal Ruiz, por sus acertadas sugerencias y al Dr.C. Arquímedes Montoya Pedrón, por el apoyo y confianza generada.

Al profesor Miguel Toledo González por la revisión de la redacción y la corrección del estilo. A mis compañeros de trabajo, en especial a los profesores Dra. Yamila Leyva, Dra. Ania Pupo, Licenciada Lourdes Almaguer por su ayuda incondicional.

A todos los que me apoyaron en esta obra, que son muchas personas...mi infinito agradecimiento.

MUCHAS GRACIAS

DEDICATORIA

A mi madre, por guiarnos a mi hermano y a mí, por un camino de sabiduría, trabajo y amor en la vida.

A la Revolución donde me formé como médico y a Fidel, soldado de las ideas, pensador con visión de futuro, creador del programa del médico y enfermera de la familia, una visión adelantada y la expresión más acabada de lo que deben ser los cimientos de la atención primaria de salud en Cuba y en el mundo.

A Sheila María, Shaila Lorena, Sergio Luis, Christina de la Caridad, Daniel y Brian Antonio, (...) pues son el futuro (...) en la medicina, las artes, la programación informática y la inteligencia artificial.

A Luisi y Mayra por su apoyo, ayuda y cariño infinito.

A Yaumara por su comprensión, amor y cariño.

A mi patria, Cuba, ese pequeño, pero gigante Caimán Verde.

SÍNTESIS

Introducción. Las enfermedades cardiovasculares son un problema de salud creciente, en Cuba no hay modelos predictivos para hacer estimaciones del riesgo cardiovascular basados en cohortes de su población. **Objetivo.** Diseñar un modelo predictivo mediante técnicas de inteligencia artificial para predecir el riesgo cardiovascular en la atención primaria de salud. **Método.** La fuente de datos fue una cohorte prospectiva seguida durante 10 años, en el policlínico José Ávila Serrano, de Velasco, Holguín. Fueron utilizadas la metodología KDD y la herramienta de minería de datos Weka, que permitieron identificar predictores significativos, para generar los modelos se aplicaron algoritmos de machine learning, se compararon los modelos obtenidos aplicándose las métricas más usadas para clases desbalanceadas. **Resultados.** Los algoritmos JRip, J48 y MLP fueron efectivos para generar los modelos, el mejor desempeño fue con el Multilayer Perceptron, una red neuronal artificial, el modelo fue implementado basado en el algoritmo de árboles de decisión J48 y el de reglas JRip y resultó más efectivo al compararlo con las tablas de predicción de riesgo cardiovascular de la OMS 2019 válidas para su uso en la población cubana. **Conclusiones.** El modelo predictivo generado mediante técnicas de inteligencia artificial fue útil, válido y comprensible, y su implementación en la atención primaria de salud supone una contribución a la prevención y control de las enfermedades cardiovasculares y evidenció un rendimiento superior al compararlo con las tablas predictivas de la OMS 2019.

TABLA DE CONTENIDOS

Contenido	Pág.
INTRODUCCIÓN	1
CAPÍTULO 1. ENFERMEDADES CARDIOVASCULARES, RIESGO CARDIOVASCULAR E INTELIGENCIA ARTIFICIAL	11
1.1 Enfermedad cardiovascular	11
1.1.1 Magnitud de la enfermedad cardiovascular en el mundo y Cuba	12
1.1.2 Aterosclerosis como proceso subyacente de las enfermedades Cardiovasculares	14
1.1.3 Factores de riesgo de enfermedad cardiovascular	20
1.2 Riesgo cardiovascular	24
1.3 Inteligencia artificial	32
1.4 Conclusiones del capítulo 1	36
CAPÍTULO 2. DISEÑO METODOLÓGICO	38
2.1 Diseño general de la investigación	38
2.1.1 Estudio de cohorte prospectivo PredRCG_aps	38
2.1.2 Diseño y validación del modelo predictivo. Metodología KDD	43
2.1.3 Comparación del modelo predictivo con las Tablas de Predicción del Riesgo de la OMS 2019	49
2.2 Métodos empleados en la investigación	50
2.3 Análisis estadístico	52
2.4 Ética	53
Consideraciones finales del capítulo 2	53

CAPÍTULO 3. RESULTADOS Y DISCUSIÓN	55
3.1 Resultados y discusión	55
3.1.1 Estudio de cohorte prospectivo	55
3.1.2 Diseño y validación del modelo predictivo. Metodología KDD	56
3.1.3 Comparación del modelo predictivo con Tablas de la OMS	69
3.2 Rol de los nuevos predictores en la aterosclerosis	78
3.3 Implicaciones del modelo en la atención primaria de salud	89
3.4 Conclusiones del capítulo 3	97
CONCLUSIONES	99
RECOMENDACIONES	100
REFERENCIAS BIBLIOGRÁFICAS	101
ANEXOS	

INTRODUCCIÓN

Las enfermedades cardiovasculares (ECV) son las enfermedades crónicas no trasmisibles más comunes, responsables de aproximadamente 17 millones de muertes a nivel mundial, estas afectan de manera creciente, a poblaciones en edad laboral y contribuyen desproporcionadamente a la pérdida de años potenciales de vida saludable y de productividad económica, situación que es reconocida como un problema de salud pública creciente en el mundo y en Cuba.^{1,2,3}

En el año 2012 las enfermedades del corazón ocuparon la segunda causa de mortalidad en Cuba, con una tasa de 197,6 por 100 000 habitantes, el 69 % de las muertes por enfermedades del corazón se produjo por enfermedades isquémicas, de ellas, un 42 % por infarto agudo del miocardio.⁴

De igual forma, en el año 2020, las enfermedades del corazón, ocuparon el primer lugar dentro de las causas de muerte con una tasa de 267,3 por 100 000 habitantes, incrementándose también la mortalidad por enfermedades cerebrovasculares.⁵

La estimación de la probabilidad de que un individuo desarrolle un evento cardiovascular a partir de sus factores de riesgo constituye un instrumento valioso para la prevención cardiovascular.¹

Décadas atrás, los investigadores del estudio de *Framingham* acuñaron el término “factor de riesgo”, que constituye un factor predictivo independiente y significativo del riesgo de sufrir la enfermedad de interés, como las cardiovasculares.⁶

Las tablas de riesgo cardiovascular son métodos simplificados que modelan la estimación del riesgo de los individuos de distintas cohortes de poblaciones seguidas generalmente

durante 10 años, en las que se establece un algoritmo matemático que permite obtener el porcentaje de sujetos que pueden desarrollar un evento cardiovascular.⁷

Han sido desarrollados, en diferentes países, varios modelos predictivos tradicionales (M-Ts), que usualmente estiman el riesgo para un periodo de 10 años, usando factores de riesgo clásicos, como Framingham clásica y por categorías, REGICOR (Registro Gironí del Cor) para España, SCORE (*Systematic COronary Risk Evaluation*), PROCAM (*PROspective Cardiovascular Münster*) y reciente las tablas de la Organización Mundial de la Salud / Sociedad Internacional de Hipertensión (OMS/SIH) para diferentes regiones del mundo.⁸

La prevención cardiovascular basada en una evaluación del riesgo cardiovascular global permite tomar decisiones más eficientes que mediante el abordaje de sus componentes aislados, sin embargo, los factores de riesgo no se distribuyen homogéneamente ni tienen el mismo efecto en todas las poblaciones.⁹

A pesar de la disponibilidad de múltiples herramientas y M-Ts, se ha observado que estos enfoques presentan limitaciones teórico metodológicas significativas en su capacidad para predecir con fiabilidad el riesgo cardiovascular.

La limitación más importante de los M-Ts es su baja sensibilidad, de manera que gran parte de los acontecimientos cardiovasculares se presentan en el grupo de la población clasificado como riesgo bajo o intermedio, otras limitaciones: radican en la necesidad de recalibrarse cuando se usan directamente en poblaciones distintas de donde fueron diseñados; consideran un número limitado de factores de riesgo; no incluyen otros factores metabólicos y biomarcadores de inflamación, a pesar de que estos acarrear riesgo independiente de complicaciones vasculares; dan un estimado de la probabilidad de que el

acontecimiento de interés ocurra en un tiempo definido, pero no establecen per se niveles de riesgo y los puntos de corte se establecen generalmente por consenso.^{1,7,8,9,10,11}

En la población cubana se han usado diferentes M-Ts, incluidas las tablas de la Organización Mundial de la Salud (OMS), las válidas para Cuba, que estiman como riesgo bajo o intermedio un gran porcentaje de la población, lo que no se corresponde con la realidad clínica de la morbilidad y mortalidad de la enfermedad cardiovascular en Cuba.¹²

En Cuba se implementa el paquete técnico: “HEARTS en las Américas: Marco de evaluación para la mejora continua de la calidad en los centros de atención primaria de salud”, que es una intervención multicomponente y basada en la evidencia que utiliza el abordaje clínico de la hipertensión y del riesgo cardiovascular como puerta de entrada para mejorar la gestión integrada e integral de la prevención y el tratamiento de las ECV en los servicios de la atención primaria de salud.¹³

Se proporciona un abordaje estratégico con el objetivo de mejorar la salud cardiovascular a nivel de la atención primaria de salud para el manejo de las enfermedades cardiovasculares basado en el riesgo cardiovascular. En el año 2007, la OMS publicó las directrices para la evaluación y el manejo del riesgo cardiovascular que se usaron como marco de referencia para la elaboración de guías de práctica clínica nacionales acerca de la prevención de las ECV. También actualizó las tablas de riesgo cardiovascular en el 2019^{1,12} y, por tanto, son las recomendadas para estimar el riesgo de enfermedad cardiovascular en la población cubana.¹⁴

En efecto, se ha esperado que la predicción del riesgo cardiovascular sea una forma eficiente de alcanzar los objetivos de la OMS de reducir la mortalidad relacionada con ECV

en un 25 % para el año 2025, con beneficios significativos a través de la mejora de la eficacia de la intervención de vida y la reducción de la carga económica.³

La inteligencia artificial (*Artificial Intelligence*, AI en inglés) puede ser utilizada para abordar las limitaciones de los M-Ts y mejorar la precisión predictiva, y es el área científica de la informática encargada de la creación de algoritmos y mecanismos automatizados, que, mediante su implantación en un sistema, tiene como objetivo optimizar el rendimiento de un proceso o aumentar las posibilidades de éxito de una tarea.

La inteligencia artificial (IA) abarca el aprendizaje automático (*Machine Learning*, ML) cuyo objetivo final es diseñar algoritmos que realizan el aprendizaje en los datos disponibles, y mediante, la minería de datos (*Data Mining*, DM) transformarlos en conocimiento, lo que permite tomar decisiones o acciones inteligentes, y el aprendizaje profundo (*Deep Learning*, DL) un subconjunto del machine learning, que incluye las redes neuronales artificiales con muchas capas.^{15,16,17}

La inteligencia artificial, mediante los algoritmos ML, tiene la habilidad de operar con cantidades ingentes de datos en tiempo real y procesarlos sin importar la estructura de estos, además, cabe destacar que este proceso se lleva a cabo de una forma más rápida que la humana, siendo además menos propenso a cometer errores.^{15,16,17}

Usando técnicas estadísticas de modelización, *machine learning* y minería de datos se pueden construir modelos de analítica predictiva, que es un área de esta última y estos modelos son un componente importante en la prevención y el control de las enfermedades cardiovasculares.^{15,16}

Varios estudios muestran que los modelos de predicción cardiovascular diseñados con inteligencia artificial (M-IAs) exhiben un desempeño superior cuando son comparados con los modelos tradicionales.^{3,18,19,20}

Cuba, la provincia de Holguín y la población atendida por el policlínico José Ávila Serrano, en Velasco, donde se desarrolló la investigación, no escapan de la magnitud creciente de la morbilidad y mortalidad por enfermedades cardiovasculares desde hace varias décadas, siendo las enfermedades del corazón, quienes ocupan el primer lugar dentro de las causas de muerte.²¹

En la revisión bibliográfica realizada, no se encontraron, en la atención primaria de salud en Cuba, modelos de predicción del riesgo de enfermedad cardiovascular basados en datos de seguimiento de cohortes en poblaciones cubanas, ni diseñados con técnicas de inteligencia artificial.

Justificación de la investigación

En el mundo y en Cuba la carga de las enfermedades cardiovasculares, en el orden económico, social y humano, es un problema creciente, que afecta cada día a pacientes más jóvenes y aporta una cantidad considerable de años de vida potenciales perdidos y vividos con discapacidad,¹⁸ por tanto, urge una predicción efectiva del riesgo de enfermedad cardiovascular para contribuir a reducir esa carga creciente.

Esta investigación pretende dar respuesta a las insuficiencias en la predicción del riesgo de enfermedad cardiovascular en la atención primaria de salud, que es la puerta de entrada de nuestro sistema de salud y hacer posible, por el equipo básico de salud, una identificación más precisa de la población asintomática, pero con riesgo de enfermedad

cardiovascular y poder aplicar las intervenciones preventivas oportunamente para reducir este peligro.

Legalmente el estudio se justifica por ser una línea priorizada de investigación del Ministerio de Salud Pública. Todo lo anteriormente expuesto sería posible si la investigación pudiera dar respuesta al siguiente **problema científico**:

Los modelos predictivos tradicionales disponibles para la evaluación del riesgo de enfermedad cardiovascular presentan limitaciones teórico metodológicas significativas en su capacidad para predecir con fiabilidad el riesgo de enfermedad cardiovascular, por tanto, se propone desarrollar un modelo predictivo basado en inteligencia artificial que mejore la precisión en la predicción y supere esas limitaciones.

Se declara como **objeto de la investigación** a los factores de riesgo cardiovascular, como **campo de la investigación** a un análisis predictivo sobre el riesgo de enfermedad cardiovascular.

Hipótesis. Un modelo predictivo que incluya factores de riesgo cardiovascular clásicos y emergentes basado en técnicas de inteligencia artificial y su implementación en la práctica médica, pudiera favorecer una predicción fiable del riesgo de enfermedad cardiovascular en pacientes asintomáticos en la atención primaria de salud.

Para dar respuesta al problema científico e hipótesis planteada, se formularon los siguientes objetivos:

1. Diseñar y validar un modelo predictivo para predecir el riesgo de enfermedad cardiovascular en la atención primaria de salud.

2. Comparar el modelo predictivo implementado con las tablas de predicción de la Organización Mundial de la Salud de 2019, válidas para su uso en la población cubana.

El **aporte teórico** de la presente investigación está dado en un modelo de analítica predictiva, como concepción teórica e integradora para la predicción precisa del riesgo de enfermedad cardiovascular, basado en técnicas de inteligencia artificial, que integra factores emergentes tomados en cuenta para su diseño, que resultan significativos en la fisiopatología de la aterosclerosis y no son incluidos en los modelos de predicción actuales, por lo que el modelo diseñado con técnicas de IA, podría ser más efectivo en la predicción del riesgo de enfermedad cardiovascular que los modelos predictivos existentes.

El **aporte práctico** radica en que se incluyen en el modelo predictivo con inteligencia artificial (M-IA) predictores de riesgo noveles, estrechamente relacionados con la aterosclerosis, que es el proceso subyacente de las enfermedades cardiovasculares, como: la proteína c reactiva de alta sensibilidad, que permite identificar la inflamación crónica o de bajo grado y que el autor la considera más que un marcador de riesgo, como un predictor de riesgo, un objetivo o blanco terapéutico en el control del riesgo cardiovascular; de igual forma la glucemia alterada en ayunas, la obesidad abdominal con la medición de la circunferencia abdominal y los triglicéridos (TG), todos predictores no considerados en los M-Ts disponibles, ni en el recomendado su uso para la población cubana según el paquete técnico HEARTS¹³, que pudieran mejorar la predicción del riesgo de enfermedad cardiovascular en la atención primaria de salud y permitiría al médico de familia constar con un M-IA robusto y accesible para una predicción efectiva del riesgo cardiovascular y así

poder aplicar oportunamente medidas preventivas efectivas para la disminución de la morbilidad y mortalidad cardiovascular.

Otro aporte es la clasificación del riesgo en solo dos categorías, estableciendo un nivel de riesgo binario, eliminando la categoría de riesgo moderado, el autor considera nombrarlo intermedio para resaltar que está entre los extremos, categoría que actualmente necesitaría otras investigaciones, no siempre disponibles en la atención primaria, para reclasificar el riesgo, y así evitar la “inercia terapéutica” de no ofrecer la intervención adecuada a su población en la categoría de riesgo intermedio.

En cuanto a la predicción del riesgo en solo dos categorías, permite una estrategia de intervención personalizada según la situación de cada paciente. El conocimiento estos de los factores de riesgo cardiovascular, le permitiría una mayor motivación y adherencia al tratamiento para modificar su nivel de riesgo y el M-IA implementado permitiría contribuir al campo de la medicina preventiva y personalizada en la atención primaria de salud, ofreciendo una herramienta eficaz para la identificación temprana de pacientes en riesgo y al médico de familia, aplicar una medicina de precisión e individualizada que permita decidir cuándo y cómo brindar atención médica de calidad.

Es necesario resaltar, que no solo clasifica las categorías de riesgo, sino aporta también, un abordaje efectivo para el diagnóstico y control de la dislipemia aterogénica, el síndrome metabólico y, por consiguiente, de la aterosclerosis subclínica, permite, además, identificar los indicadores de carga alostática, la meta inflamación de bajo grado y sospechar la resistencia a la insulina en la atención primaria de salud.

La **novedad científica** consiste en que se presenta a la comunidad científica especializada un modelo predictivo moderno de enfermedad cardiovascular, con datos provenientes de

un estudio de cohorte de una fracción de una población cubana y diseñado con técnicas de inteligencia artificial, que supera las limitaciones de los modelos predictivos disponibles, además, incluye factores de riesgo clásicos y emergentes, que al ser implementado en la atención primaria de salud podría favorecer una predicción eficaz del riesgo cardiovascular y permitiría aplicar medidas preventivas oportunamente para el control de los factores de riesgo cardiovascular y por consiguiente, una disminución de la morbilidad y mortalidad por enfermedades cardiovasculares.

Como **significación práctica**, la implementación del modelo predictivo, le permite al médico de la atención primaria de salud, disponer de un instrumento efectivo, comprensible y fácil de usar para la predicción del riesgo de enfermedad cardiovascular, que supera las limitaciones de los M-Ts y mejora la precisión predictiva al realizar una evaluación más completa del riesgo cardiovascular, que contribuiría a cerrar la brecha actual en la predicción y reducir la carga de las enfermedades cardiovasculares, de ahí su impacto preventivo relevante en la atención primaria de salud.

Con la consideración del problema de investigación, los objetivos e hipótesis la investigación fue realizada en tres etapas estrechamente relacionadas: el ensamble y seguimiento del estudio de cohorte prospectivo que fue la fuente de los datos; el diseño y validación del modelo predictivo, usando la metodología de Descubrimiento de Conocimientos en Bases de Datos, *Knowledge Discovery in Databases*, conocido como **KDD**, por sus siglas en inglés, empleando la herramienta de minería de datos Weka y finalmente, en la tercera etapa fue comparado el modelo implementado con las tablas predictivas de la Organización Mundial de la Salud para las Américas, región del Caribe,

recalibradas en el 2019 y válidas para su uso en la población cubana, recomendadas por el paquete técnico HEARTS.¹³

Se utilizaron métodos empíricos, teóricos y estadísticos, que se explican y detallan en el capítulo correspondiente, entre los cuales se encuentran: análisis documental, método histórico-lógico; método hipotético-deductivo; método de análisis y síntesis; modelación, método holístico-dialéctico; método estadístico-analítico.

El procesamiento estadístico se realizó con el *Statistical Package for the Social Sciences* (SPSS) versión 25.0 para *Windows*, el *software* Epidat 4,2 y con el *software* de minería de datos Weka, versión 3.9.5.

La tesis está estructurada de la siguiente manera: introducción, tres capítulos, conclusiones, recomendaciones, referencias bibliográficas y anexos.

El primer capítulo relata la magnitud de las enfermedades cardiovasculares, la aterosclerosis como proceso subyacente de las mismas; las funciones de riesgo cardiovascular, su fundamento y limitaciones, las funciones predictivas utilizadas en la población cubana y sus limitaciones, así como, la inteligencia artificial, el aprendizaje automático, la minería de datos y el análisis predictivo.

El segundo capítulo detalla el diseño metodológico utilizado y en el tercero se analizan y discuten los resultados de la investigación.

CAPÍTULO I. ENFERMEDAD CARDIOVASCULAR, RIESGO CARDIOVASCULAR E INTELIGENCIA ARTIFICIAL

En este capítulo se expone el estado de la temática, que es la base teórica para abordar la investigación, incluye la magnitud de las enfermedades cardiovasculares, la aterosclerosis como proceso subyacente de las mismas, los factores y las funciones de riesgo cardiovasculares, el fundamento teórico de su cálculo y sus limitaciones, las funciones predictivas aplicadas a la población cubana, sus resultados, así como, la inteligencia artificial, el *machine learning*, la minería de datos y el análisis predictivo.

Objetivos:

1. Describir la magnitud del problema de las enfermedades cardiovasculares, dentro de las enfermedades no transmisibles, a nivel mundial, en Las Américas y en Cuba, así como el rol que desempeña la aterosclerosis como proceso subyacente de estas enfermedades y la importancia del control de los factores de riesgo de enfermedad cardiovascular.
2. Exponer acerca del fundamento teórico del cálculo del riesgo cardiovascular, las principales funciones de riesgo, las funciones predictivas aplicadas a la población cubana y sus limitaciones.
3. Explicar sobre la inteligencia artificial, el *machine learning*, la minería de datos y el análisis predictivo

1.1 Enfermedad cardiovascular

1.1.1 Magnitud del problema de las enfermedades cardiovasculares, dentro de las enfermedades no transmisibles, a nivel mundial, en Las Américas y en Cuba

Las enfermedades no transmisibles (ENT), suelen ser de larga duración y son el resultado de una combinación de factores genéticos, fisiológicos, ambientales y de comportamiento, las principales ENT son las enfermedades cardiovasculares, el cáncer, las enfermedades respiratorias crónicas y la diabetes mellitus.²²

Las enfermedades no transmisibles se cobran la vida de 41 millones de personas anualmente, 17 millones de personas mueren por una ENT antes de los 70 años de edad, las ECV representan la mayor parte de la mortalidad por ENT, es decir, 17,9 millones de personas al año.²²

Las ECV son la principal causa de morbilidad y muerte en todo el mundo, contribuyendo a gastos enormes en los sistemas de salud pública²³, cada año mueren más personas por ECV que por cualquier otra causa, se calcula que en 2015 murieron por esta causa 17,7 millones de personas, que representa un 31 % de todas las muertes registradas en el mundo, de estas muertes, 7,4 millones se debieron a la cardiopatía coronaria y 6,7 millones, a los accidentes cerebrovasculares.²⁴

Las muertes por ECV se incrementaron de 12,1 millones en 1990 a 18,6 millones en 2019 y se estima alcancen los 24 millones en 2030²³, en los accidentes cerebrovasculares se espera un aumento de 5 millones para 2030, por lo que es imperativo el control de los factores de riesgo cardiovascular (FRCV).²⁵

Debido a que, al menos tres cuartas partes de las defunciones causadas por ECV en el mundo se producen en los países de ingresos bajos y medios, muchos habitantes de dichos países mueren más jóvenes, de ordinario en la edad más productiva, además suponen una pesada carga económica, social y humana²⁴.

Las ECV siguen siendo la causa principal de mortalidad, carga de enfermedad y discapacidad en la Región de las Américas, en 2019, 2 millones de personas murieron a causa de las enfermedades cardiovasculares, con 36.4 millones de años de vida perdidos por muerte prematura, 4.5 millones de años vividos con discapacidad, el número de años de vida vividos con discapacidad casi se duplica en las dos últimas décadas, la cardiopatía isquémica y el accidente cerebrovascular fueron las dos principales causas de mortalidad y discapacidad.²⁶

La tendencia de la mortalidad y carga de las enfermedades cardiovasculares estandarizadas por edad y en ambos sexos en la Región de las Américas desde el 2000 al 2019, que es la última información disponible, se muestra en el gráfico 1 (Anexo 1), fue utilizado el método de la clasificación de quintiles, por lo que cada intervalo contiene el 20% de los países, con una mediana, en la región de 150,7 por 100 000 habitantes.²⁷

El Caribe es una de las subregiones del mundo con mayor prevalencia de ENT, en primer lugar, las enfermedades cardiovasculares, constituyen la principal causa de muertes en los países de la subregión, siendo una fuente significativa de discapacidad y en la mayoría de los países, su control no está enfocado suficientemente en la prevención, ni están teniendo enfoques multisectoriales y coordinados.²⁸

Algo semejante ocurre en Cuba, las principales causas de muerte, en los últimos 15 años, desde 2009 al 2018 han sido la cardiopatía isquémica y las enfermedades cerebrovasculares, que se sitúan como la primera y tercera causa de fallecimiento, las tasas de mortalidad por enfermedades del corazón en el periodo se incrementaron en todas las provincias cubanas.²⁹

En Cuba en el año 2018, 25 mil 684 personas murieron de una enfermedad del corazón lo que representa una tasa de 228,2 x 100 000 habitantes, de ellos 16 260 individuos fallecieron por una enfermedad isquémica cardiaca, para una tasa 144,5 x 100 000 habitantes, de estos fallecidos 7 350 tuvieron como principal causa de muerte el infarto agudo del miocardio, con una tasa de 65,3 x 100 000 habitantes.³⁰

En el año 2021, las enfermedades del corazón, se mantienen como la primera causa de muerte con una tasa de 384,9 por 100 000 habitantes, el 60,0 % de estas por enfermedades isquémicas; y el 37,4 % por infarto agudo del miocardio.^{31,32}

Por otra parte, la tasa promedio de mortalidad por infarto agudo del miocardio en la población de 40 a 79 años, en Cuba, desde el año 2012 al año 2018, utilizando la población y las defunciones, con los códigos 120 y 121 de la Clasificación Internacional de Enfermedades (CIE-10), se manifestó de la siguiente manera: en los años 2012, 2014, 2016 y 2018 fue de 11,4 x10 000 habitantes en el sexo masculino y 5,9 x10 000 en el femenino, mostrando (Anexo 1, gráfico 2) una tendencia ligeramente ascendente en el sexo masculino y se mantuvo sin cambios en el femenino.^{4,30,33,34}

En síntesis, actualmente las enfermedades cardiovasculares crean un importante problema de salud en Cuba y en la mayoría de los países de las Américas y del mundo.

1.1.2 Aterosclerosis como proceso subyacente de las enfermedades cardiovasculares

Etiopatogenia de aterosclerosis

El problema subyacente de las enfermedades cardiovasculares, es la aterosclerosis, los episodios coronarios y los accidentes cerebrovasculares agudos, en su mayoría, se

producen de forma repentina y conducen a menudo a la muerte antes de que pueda dispensarse la atención médica requerida.³⁵

La aterosclerosis es una enfermedad vascular crónica que afecta las arterias de mediano y gran calibre. Comienza en la vida fetal, progresa lentamente durante la niñez y la adolescencia, y su desarrollo se acelera en la vida adulta, la inflamación y el estrés oxidativo desempeñan una función importante en cada una de sus etapas.^{36,37}

La fisiopatología de la aterosclerosis interrelaciona procesos moleculares y celulares complejos, transitando por varias etapas que van desde los procesos subclínicos más tempranos, como la disfunción endotelial y formación de estrías grasas, hasta estadios más avanzados con el engrosamiento de la íntima, fibrosis y remodelación del tejido, acumulación de restos necróticos, de células T, macrófagos y lipoproteínas aterogénicas oxidadas, calcificación y finalmente la ruptura de la placa de ateroma.³⁷

A finales de los años setenta, se destacó la importancia de la proliferación de células de músculo liso en la formación de lesiones ateroscleróticas y la hipótesis acerca de que la lesión de la pared arterial tiene un papel preponderante en la progresión de la placa, posteriormente, se identificó la inflamación como la principal fuerza impulsora de la activación y proliferación de células de músculo liso, procesos mediados por factores de crecimiento, estudios subsecuentes a finales de los años noventa, indicaron que el mecanismo de progresión de la enfermedad implicaba una interacción compleja entre los factores de riesgo y la inflamación.^{37,38}

Aterosclerosis subclínica y dislipemia aterogénica

Como se ha mencionado las lesiones ateroscleróticas se van depositando en silencio día por día en las arterias y en ocasiones su primer síntoma es un evento cardiovascular agudo

mortal o no, cuando existe un grado de evidencia de lesión de la pared arterial sin expresión clínica usamos el término de aterosclerosis subclínica.³⁹

La aterosclerosis subclínica (AS) se define como la presencia de cualquier placa de ateroma en territorio carotideo, aórtico o iliofemoral, o por la presencia de una puntuación de calcificación arterial coronaria (CACs) mayor de 1,³⁷ se estima que alrededor del 40-60% de los episodios isquémicos de origen aterotrombótico ocurren como una primera manifestación, sin signos previos que puedan alertar de estos.⁴⁰

Por lo tanto, la prevención temprana es el mejor enfoque para combatir esta enfermedad pandémica, que tiene una fase subclínica prolongada, y se suele diagnosticar en etapas avanzadas o ante un evento cardiovascular, su identificación ha demostrado que mejora la estratificación del riesgo con respecto a los métodos convencionales.⁴¹

Prueba de ello, el estudio PESA (*Progression Early Subclinical Atherosclerosis*) evaluó la presencia de aterosclerosis mediante técnicas no invasivas en personas asintomáticas, se trató de un estudio prospectivo de cohortes en más de 4.000 individuos asintomáticos de 40 a 54 años, donde se evaluó el riesgo cardiovascular mediante distintos sistemas: Framingham, SCORE europeo y el nuevo algoritmo de la AHA (*American Heart Association*), se categorizó a los pacientes en riesgo bajo, intermedio y alto, además fue realizada la detección de aterosclerosis mediante score de calcio coronario (SCC) y ultrasonido en los territorios carotideos, aorta abdominal e iliofemorales. Para analizar la extensión de la enfermedad, se categorizó la aterosclerosis como ausente, focal (1 territorio), intermedia (2-3 territorios) y generalizada (4-6 territorios).⁴¹

Finalmente, en un estudio de 498 sujetos asintomáticos que fueron explorados mediante medición del calcio coronario y aórtico con tomografía computarizada (TC) y mediante

ecografía carotídea, a los que se les aplicó la ecuación de riesgo de Framingham, 312 pacientes (63 %) fueron considerados de bajo riesgo y de estos 214 (69 %) tenían evidencia de aterosclerosis subclínica en al menos un lecho vascular (coronario, aórtico o carotideo), asimismo, en otro análisis retrospectivo de 222 adultos que habían sufrido un infarto del miocardio prematuro, solo un 27 % de los pacientes fueron categorizados como de alto riesgo antes del infarto, mientras que el 50 % fue clasificado como de bajo riesgo y solo el 50 % tenía uno o ningún factor de riesgo cardiovascular clásico antes del infarto agudo del miocardio.⁴¹ A consideración del autor esos M-Ts evidencian las limitaciones para una predicción precisa del riesgo de enfermedad cardiovascular, ya que solo incluyen como predictores factores de riesgo clásicos o tradicionales.

Al mismo tiempo, las técnicas de imagen no invasivas han permitido evaluar la presencia de aterosclerosis en pacientes asintomáticos, un análisis multiterritorial tiene el potencial de brindar una visión integral de la distribución y carga de aterosclerosis.⁴⁰

Actualmente, existe mucho interés en desarrollar otras técnicas proteómicas de última generación, rápidas y no invasivas para detectar la carga aterosclerótica, la identificación de biomarcadores plasmáticos estrechamente asociados con la aterosclerosis subclínica, podría superar las limitaciones de los modelos predictivos disponibles y mejorar la predicción del riesgo cardiovascular.⁴²

Dislipemia aterogénica

La dislipemia aterogénica (DA) es una tríada caracterizada por niveles elevados de triglicéridos, niveles normales o elevados de colesterol unido a lipoproteína de baja densidad (cLDL), que son pequeñas y densas, altamente aterogénicas, con valores bajos de colesterol unido a lipoproteínas de alta densidad (cHDL), esta condición lipídica es

favorecida por un estado de resistencia insulínica persistente, que permite la liberación de ácidos grasos libres desde el tejido adiposo abdominal, el cual es endocrinológicamente activo; esos ácidos grasos viajan al hígado donde forman parte de las lipoproteínas de muy baja densidad (VLDL), ricas en triglicéridos; estas salen del hígado e intercambian triglicéridos por colesterol esterificado con las otras lipoproteínas circulantes. Este perfil lipídico alterado favorece la iniciación y desarrollo acelerado de la aterosclerosis, clínicamente asociada a las ECV.^{43,44}

Igualmente, la DA, se asocia a enfermedades muy prevalentes en la población con alto riesgo cardiovascular como: el sobrepeso, la obesidad, la diabetes mellitus tipo 2 y el síndrome metabólico, por consiguiente, la DA es por sí misma un indicador de alto riesgo cardiovascular.⁴⁵

La importancia de la dislipemia aterogénica, muy relacionada con la aterosclerosis subclínica, viene dada por ser una causa fundamental en las muertes debidas a enfermedades cardiovasculares, que son a su vez la primera causa de mortalidad en la mayoría de los países del mundo y, en Cuba, sus componentes están bien identificados y son fácilmente reconocibles, sin embargo, esta dislipemia generalmente, esta infradiagnosticada, infratratada y en consecuencia infracontrolada.⁴⁶

Síndrome metabólico

El síndrome metabólico es un mosaico de problemas que dañan seriamente la salud y su forma de presentación es variable en un mismo individuo; unido a esto, los factores tanto genéticos como ambientales concommitan con el estilo de vida y la resistencia a la insulina (RI), que es el factor patogénico fundamental.⁴⁷

A su vez, las personas con síndrome metabólico tienen de dos a tres veces el riesgo de desarrollar enfermedad cardiovascular aterosclerótica y un riesgo cinco veces mayor de desarrollar diabetes mellitus tipo 2 (DM2) y muerte por todas las causas.^{45,46}

El síndrome metabólico (SM) es una combinación de desórdenes médicos y se ha convertido en un serio problema de salud en todo el mundo.⁴⁷

De ahí que, la importancia clínica del SM se deriva de dos hechos; en primer lugar, de su capacidad para predecir la diabetes mellitus tipo 2, la mortalidad por enfermedad cardiovascular y por otras causas y, en segundo lugar, de su elevada prevalencia.⁴⁷

La prevalencia del SM varía, pero el consenso más generalizado coincide en que, independiente de los criterios diagnósticos que se utilicen, cerca del 20 al 40 % de la población adulta se ve afectada por este, la relevancia clínica del SM se relaciona también con su presencia en el 80 a 85 % en los pacientes con diabetes mellitus tipo 2 y con las posibles complicaciones relacionadas, que son progresivas y acumulativas en dependencia del tiempo de evolución.⁴⁸

Es preciso insistir que los criterios diagnósticos de SM tienen limitaciones en su utilidad clínica, pues no incluyen otros factores de riesgo, predictores de diabetes y enfermedad cardiovascular como: edad, sexo, historia familiar de diabetes o enfermedad cardiovascular, condiciones socioeconómicas, cHDL, proteína c reactiva e importantes variables de conducta como consumo de tabaco y actividad física.^{48,49}

Por ende, si la prevalencia del SM en la población adulta es aproximadamente del 20 a 40 % y los componentes lipídicos de dicho síndrome coinciden con los de la DA, se puede estimar la gran prevalencia de este trastorno en la población general.^{49,50}

Los M-Ts evalúan el riesgo de enfermedad cardiovascular analizando una serie de factores de riesgo cardiovasculares clásicos como el tabaquismo, el sobrepeso o la hipertensión, sin embargo, muchos individuos que no tienen ningún factor de riesgo desarrollan placas ateroscleróticas; es por ello que la detección temprana de la aterosclerosis subclínica, debido a su alta prevalencia en la población general, reviste un gran interés biomédico.^{49,50}

Por consiguiente, están plenamente identificados los criterios diagnósticos de la dislipemia aterogénica, el síndrome metabólico y su evidente relación con la aterosclerosis subclínica, estas entidades comparten el mismo perfil metabólico y una inmuno inflamación crónica o de bajo grado. Se aprecia un incremento de su prevalencia, incluso en edades más jóvenes asociados a la obesidad y a la DM2, pero en la atención primaria de salud, están infradiagnosticadas y, por consiguiente, no controladas, una situación similar es reportada en Latino América.^{51,52,53}

En opinión del autor, se debe mejorar en la atención primaria de salud, el diagnóstico oportuno y, por consiguiente, el control de estas entidades muy relacionadas con la aterosclerosis y lograr la prevención de las enfermedades cardiovasculares.

1.1.3 Factores de riesgo de enfermedades cardiovasculares

Clasificación de los factores de riesgo cardiovascular

En la actualidad, se evidencia un avance de la medicina cardiovascular, la implementación de nuevas técnicas de imagen y herramientas terapéuticas, permiten acercarse a la enfermedad cardiovascular desde una perspectiva más holística, donde se tiende a potenciar la prevención y el diagnóstico precoz e, identificar la enfermedad en sus fases iniciales y lograr el control de los FRCV, tanto clásicos como emergentes.⁵⁴

El concepto de factor de riesgo fue introducido en el estudio Framingham y desde entonces los estudios epidemiológicos en grandes poblaciones han permitido identificar una serie de factores de riesgo de enfermedad cardiovascular, que actúan como inductores en la formación de la placa de ateroma.⁶

Un FRCV es una característica biológica, un hábito o estilo de vida que aumenta la probabilidad de padecer o de morir a causa de una enfermedad cardiovascular en aquellos individuos que lo presentan, precisamente, al tratarse de una probabilidad, la ausencia de los factores de riesgo no excluye la posibilidad de desarrollar una ECV en el futuro, y la presencia de ellos tampoco implica necesariamente su aparición.⁵⁴

Se considera factor de riesgo aquel que está asociado con una enfermedad porque se encuentra en la vía causal que la produce.⁵⁵

De manera que, los factores de riesgo clásicos o tradicionales son aquellos que, en estudios experimentales, epidemiológicos, clínicos, observacionales y de intervención, han mostrado ser importantes en la aparición y progresión de las enfermedades cardiovasculares, y cuyo control lleva a claros beneficios en la incidencia de las mismas.⁵⁶

Los principales factores de riesgo pueden ser no modificables (edad, sexo, factores genéticos/historia familiar) o modificables, estos precisamente los de mayor interés, ya que en ellos cabe actuar de forma preventiva, siendo muy frecuentes en la población.^{54,57}

Otros FRCV que se asocian a un mayor riesgo de ECV son el cHDL bajo y los triglicéridos altos, componentes habituales del SM y la DA, que juntos potencian un mayor riesgo cardiovascular y de desarrollar ECV y DM2.⁵⁴

Es necesario resaltar que una limitación de los M-Ts es que la mayoría utilizan como predictores factores de riesgo clásicos o tradicionales.

Factores de riesgo emergentes

En los últimos años han surgido con fuerza estudios que proponen la introducción de nuevos FRCV o emergentes, también llamados no tradicionales o noveles, con la pretensión de que sean incorporados en la evaluación y la estratificación del riesgo cardiovascular en los individuos y en las poblaciones, con las consiguientes implicaciones en las decisiones preventivas y terapéuticas.^{58,59}

Los denominados factores de riesgo emergentes, son aquellos que en estudios básicos, epidemiológicos o clínicos se muestran como posibles elementos que participan en la génesis de las ECV, pero que todavía se necesitan más evidencias para considerarlos como factores de riesgo.^{58,59}

Los factores de riesgo emergentes tienen implicaciones clínicas potencialmente importantes para la prevención cardiovascular, ya sea a través de la mejora en la predicción del riesgo o para tratar las enfermedades cardiovasculares, dado que los factores de riesgo clásicos no explican la variación interindividual del riesgo cardiovascular, pues se estima que un 10 % a 15 % de los pacientes que padecen una enfermedad cardiovascular no tienen ningún factor de riesgo cardiovascular clásico.^{57,59}

Los FRCV emergentes, según la revisión de la literatura, se muestran en la tabla 1 (Anexo 1).^{58,59}

Hoy, se concede gran importancia a los factores psicosociales, como el bajo nivel socioeconómico, el aislamiento social, la depresión u hostilidad y el estrés laboral o familiar;

ya que además de asociarse a un mayor riesgo cardiovascular, empeoran el pronóstico de los pacientes con cardiopatía isquémica establecida y dificultan significativamente el control de los FRCV clásicos.⁶⁰

Es preciso tener presente, que no fue objetivo de la investigación abordar los determinantes sociales de salud, que son factores que influyen en la fuerza de vida de las personas y las comunidades.

La proteína c reactiva de alta sensibilidad (pcr-as), es considerada un factor emergente, ya que puede predecir el riesgo cardiovascular en una amplia variedad de situaciones clínicas, por lo que su evaluación ha recibido gran atención y se ha sugerido que podría ser usada no solo como un biomarcador de riesgo del estado aterosclerótico, sino que podría considerarse también como un factor predictor de riesgo.^{58,61}

Búsqueda de nuevos marcadores de lesiones de aterosclerosis

En otro sentido, las concentraciones de moléculas que participan en los diferentes procesos patológicos presentes en la aterosclerosis podrían ser biomarcadores en teoría, los biomarcadores estudiados hasta la actualidad se han basado en la posibilidad de que sean útiles desde el punto de vista diagnóstico o pronóstico, aunque algunos no tengan valor diagnóstico y terapéutico, pueden proporcionar información sobre la génesis y la formación de la placa ateromatosa.⁶²

Un marcador de riesgo es aquel que está asociado con la enfermedad, pero que no necesita estar causalmente relacionado con ella y puede ser una medida del proceso de la enfermedad por sí mismo.⁶²

Con los avances recientes en la fisiopatología de la aterosclerosis, el vínculo entre inflamación y aterosclerosis está bien establecido, la inflamación es el elemento clave en el proceso aterosclerótico y contribuye en todos sus estadios: iniciación, crecimiento y ruptura de la placa de ateroma. Hay una serie de biomarcadores que reflejan inflamación, hemostasia, trombosis y estrés oxidativo, que han sido propuestos como herramientas clínicas potenciales en un esfuerzo por mejorar la predicción de la enfermedad cardiovascular.^{2,60,61,62}

Dicho de otro modo, un biomarcador es útil si añade evidencia clínica y proporciona información acerca del riesgo, independiente de los predictores establecidos, es fácil de medir e interpretar en la atención primaria, si se tiene en cuenta que es exacto, reproducible y está internacionalmente estandarizado.⁶²

1.2 Riesgo Cardiovascular

El riesgo cardiovascular se define como la probabilidad de padecer un evento cardiovascular en un determinado período, que habitualmente se establece en 5 ó 10 años, en los pacientes que no padecen de enfermedad cardiovascular, es decir, en prevención primaria y es fundamental para establecer la intensidad de la intervención, la necesidad de instaurar tratamiento farmacológico y la periodicidad de las visitas de seguimiento.²

La predicción del riesgo cardiovascular, ha constituido, en los últimos años, la piedra angular en las guías clínicas de prevención cardiovascular y deviene una herramienta útil para establecer prioridades en la atención primaria.^{2,63,64}

Fundamentos teóricos del cálculo del riesgo cardiovascular

Para identificar las personas con riesgo elevado en la atención primaria de salud, se determinan los factores de riesgo presentes y mediante las funciones o tablas de riesgo, se convierten en estimación del riesgo cardiovascular.^{6,7}

Existen dos métodos de cálculo del riesgo cardiovascular: cualitativos y cuantitativos. Los cualitativos se basan en la suma de factores de riesgo o la medición de su nivel; los cuantitativos, por su parte, están basados en ecuaciones de predicción de riesgo que nos dan la probabilidad de presentar un evento cardiovascular en un determinado tiempo y la forma de cálculo es a través de programas informáticos o de las llamadas tablas de riesgo cardiovascular.^{63,64,65}

Por consiguiente, las funciones de riesgo son ecuaciones que calculan la probabilidad de que un individuo presente el evento de interés, en un intervalo de tiempo, según el nivel de exposición a diferentes factores, la generación de modelos predictivos a partir de estudios de cohorte prospectivos, poseen un gran valor científico.^{66,67}

Para el cálculo de la probabilidad de aparición de un suceso dicotómico, los modelos matemáticos más habituales se basan en la utilización de la regresión logística, como el modelo de riesgos proporcionales de Cox; que es un método creador de modelos para datos de tiempo de espera hasta un evento con casos censurados presentes, que relaciona los factores predictores a la incidencia de un evento cardiovascular en un período de seguimiento, generalmente de 10 años.^{66,67,68,}

Es decir, el modelo de riesgos proporcionales de Cox, se considera como semiparamétrico debido a que incluye una parte paramétrica y otra no paramétrica, siendo un modelo bien recibido en análisis de supervivencia, como no tiene especificada la función de riesgo basal es posible estimar los coeficientes de regresión, calcular las razones de riesgo y ajustar las

curvas de supervivencia a una gran variedad de situaciones⁶⁶. En el proyecto SCORE (*Systematic Coronary Risk Evaluation*), para la población europea, fue utilizada la función de Weibull.^{67,68}

En otras palabras, los datos provienen de estudios de cohorte prospectivos, que agrupa un conjunto de personas asintomáticas que en general son homogéneas respecto a alguna característica básica y se determina la incidencia de una enfermedad cardiovascular en el tiempo, tienen la ventaja sobre los diseños transversales de separar los posibles factores de riesgo de la aparición de la enfermedad a lo largo del tiempo para evaluar las relaciones temporales.^{7,68}

Además, para el diseño de los modelos, se incluyen todas las variables predictoras, que tienen una relación monótona con la variable resultado, las variables numéricas, pueden ser transformadas en variables dummy con codificación de indicador y para determinar la estimación de la supervivencia por sexos, se utilizan las curvas de supervivencia de Kaplan-Meier.^{68,69}

Finalmente, se evalúa el desempeño del modelo, mediante la discriminación, que es la capacidad del modelo de separar a los participantes que presentaron un acontecimiento cardiovascular, que puede representarse mediante el área bajo la curva ROC (*Receiver Operating Characteristic*) y se determina la calibración, que es la precisión de la estimación del riesgo absoluto comparándola con las tasas de incidencia reales, donde se utiliza la prueba de Hosmer y Lemeshow de bondad de ajuste, basada en una distribución de Chi cuadrado.^{66,67,68,69}

Un aspecto clave es, que la valoración global del riesgo mediante modelos multifactoriales, predice el riesgo global individual de forma más exacta, en la actualidad las tablas de riesgo

se basan en algoritmos para predecir el riesgo cardiovascular y las más utilizadas están basadas en la ecuación de riesgo de Framingham.^{6,70,71}

Diferentes funciones para la estimación del riesgo cardiovascular

El Estudio Framingham del Corazón (en inglés *Framingham Heart Study*), sobre el riesgo coronario, se puso en marcha en 1948⁶⁵, para investigar una epidemia creciente de enfermedad coronaria, mediante un enfoque prospectivo, entonces novedoso para los estudios poblacionales epidemiológicos, no tardó en ponerse de manifiesto que la etiología era multifactorial; entonces se definieron los "factores de riesgo" mayores y se corrigieron algunos conceptos clínicos erróneos.⁷¹

El primer algoritmo de riesgo de Framingham fue propuesto en 1991 a partir de la primera cohorte de Framingham, los factores de riesgo considerados fueron la edad, el sexo, el colesterol total, el cHDL, la presión arterial y el tabaquismo, es actualmente el algoritmo más extendido, aunque con ajustes de calibración para diferentes países.⁶⁵

En 2008 fue desarrollado un nuevo algoritmo de Framingham, para estimar el riesgo cardiovascular global, donde se tiene en cuenta la edad, el cHDL, los TG, el colesterol total, la presión arterial sistólica con puntuaciones diferentes en función del tratamiento, el hábito de fumar y la presencia DM2.⁷¹

Muchas otras cohortes prospectivas en varios países han sido utilizadas para diseñar puntajes de riesgo, por ejemplo: REGICOR para España 2003, que propuso a partir de un estudio en la región de Girona, España, las tablas de estimación de riesgo coronario a 10 años a partir de la calibración de la ecuación de Framingham, adaptada a la población española.^{72,73}

El proyecto SCORE se inició para desarrollar un sistema de evaluación de riesgo cardiovascular para su uso en Europa, en él se calcula, mediante un modelo basado en la función de Weibull, el riesgo de enfermedad cardiovascular en 10 años, estimándose dos ecuaciones diferentes para enfermedades coronarias y no coronarias, para países de alto y bajo riesgo.^{67,74}

El puntaje PROCAM (*Prospective Cardiovascular Munster Study*), estudio prospectivo alemán que calcula el riesgo de infarto del miocardio o muerte coronaria a 10 años, utilizando como variables la edad, el cHDL, los TG, el tabaquismo, DM2, la presión arterial sistólica y los antecedentes familiares de enfermedad cardiovascular.^{72,75}

El QRISK1 y QRISK2 derivadas de una base de datos británica, que incluyó a 10 millones de pacientes, el algoritmo incluye factores de riesgo tradicionales, los antecedentes familiares, el índice de masa corporal, la toma de antihipertensivos, la raza, la artritis reumatoide, la enfermedad renal y la fibrilación auricular.^{72,76}

En el 2007, la Organización Mundial de la Salud y la Sociedad Internacional de Hipertensión (SIH) publicaron dos variantes de tablas de predicción de riesgo cardiovascular, basado en una ecuación multivariante de predicción de riesgo para cada una de las 14 subregiones epidemiológicas de la OMS: una variante con inclusión de colesterol y otra sin incluirlo, ambas tablas incorporan el sexo, la edad, el hábito de fumar, diabetes y las cifra de tensión arterial sistólica, se utilizaron datos estandarizados sobre la prevalencia de los factores de riesgo y el riesgo relativo de sufrir un evento aterotrombótico en cada una de estas regiones.⁷⁷

En 2019, la OMS para ayudar a adaptar los enfoques de predicción del riesgo de enfermedades cardiovasculares a los países de bajos y medianos ingresos, revisa los

modelos de predicción y se recalibraron utilizando las incidencias específicas de edad, sexo y los factores de riesgo disponibles en 21 subregiones globales, donde se disminuyó el umbral de riesgo. Esta tabla de la OMS 2019¹ es la que recomienda el paquete técnico HEARTS para la determinación del riesgo cardiovascular en la población cubana.^{13,14}

Los M-Ts mencionados son los más usados a nivel global para la predicción del riesgo de enfermedad cardiovascular.¹²

Estimación del riesgo cardiovascular en la población cubana

Según la literatura revisada, en Cuba se han empleado cinco diferentes M-Ts, al menos en 29 ocasiones, donde se han obtenido resultados heterogéneos, con criterios de análisis diversos y dispersos, en la inclusión de sujetos y diseño del estudio, con aportes que clasifican a la mayoría de la población como riesgo bajo o intermedio, lo que no se corresponde con la realidad clínica de la morbilidad y mortalidad por este tipo de enfermedad.¹²

Por ejemplo, en los datos de la III Encuesta Nacional de Factores de Riesgo de Enfermedades no Trasmisibles, 2010-2011, representativa de la población de 15 años y más, por sexo y grupos de edad fueron utilizadas las tablas de predicción del riesgo de la OMS de 2007. El riesgo fue clasificado en: bajo (<10 %), moderado (10-19,9 %), alto (20-29,9 %), muy alto (≥ 30 %), el 5.2 % de la población cubana tenía alto y muy alto riesgo cardiovascular, 8.6 % riesgo moderado y 86 % riesgo bajo.⁷⁷

Otra investigación sobre la epidemiología de los factores de riesgo cardiovascular y riesgo cardiovascular global se aplicó en personas de 40 a 79 años en la atención primaria, utilizando las tablas de predicción de riesgo de la OMS 2007. La estimación del riesgo cardiovascular mostró que 34,2 % de la población estudiada tenía más de 10 % de

probabilidades de sufrir un episodio coronario en los próximos 10 años, mientras que el 65,8 % tuvo un bajo riesgo.⁷⁸ En estudio realizado en una población del área de salud del Policlínico Santa Clara, usando las tablas de predicción de riesgo de la OMS 2007, el 74,6 % de la población estudiada presentó un bajo riesgo, el 16,2 % un riesgo moderado, el 6,3 % alto y el 2,8 % muy alto.⁸⁰

En el Instituto de Cardiología y Cirugía Cardiovascular de Cuba, usando la tabla de riesgo de Gaziano⁸¹ se estimó el riesgo cardiovascular en 403 trabajadores entre 35 y 70 años, el nivel de riesgo cardiovascular global alto fue de 13,4 %, en hombres 24,8 % y en mujeres 5,5 %.⁸²

En otros estudios de predicción de riesgo cardiovascular en la población cubana, utilizando las tablas de la OMS 2007 o las de Gaziano, la mayoría de los pacientes se incluyen en el nivel de riesgo bajo o moderado,^{83,84} solo se incrementa el nivel de riesgo al usar la tabla de Gaziano modificada.⁸⁵

En la revisión de la literatura no se encontraron predicciones en la población cubana utilizando las tablas de riesgo de la OMS de 2019¹, recomendado su uso por la iniciativa HEARTS.¹⁴

Limitaciones de las funciones de riesgo cardiovascular

Es importante insistir que, la mayoría de los M-Ts están basados en los criterios de predicción de la ecuación de riesgo de Framingham, que puso en perspectiva la presencia de FRCV¹² y se evidencian limitaciones teórico metodológicas en las mismas, que dificultan una predicción fiable del riesgo cardiovascular en la atención primaria de salud.

La limitación más importante de los M-Ts es su baja sensibilidad, de manera que gran parte de los acontecimientos cardiovasculares se presentan en el grupo de la población clasificado como riesgo intermedio e incluso bajo^{7,9,12}

De manera que, los M-Ts consideran un número limitado de factores de riesgo, casi en su mayoría clásicos o tradicionales, pero no incluyen otros factores metabólicos y biomarcadores de inflamación, a pesar de que estos acarrear un riesgo independiente de complicaciones vasculares^{7,9,12}

Contrariamente, no consideran que el riesgo es un continuo, geométrico más que lineal, los puntos de corte para definir el riesgo son arbitrarios y se establecen generalmente por consensos, las funciones expresan la probabilidad de que un evento ocurra en un periodo de tiempo, pero no establecen per se el nivel de riesgo.^{11,72}

Por último, los M-Ts necesitan recalibrarse cuando se utilizan en otras poblaciones, basados en la incidencia de la enfermedad y la prevalencia poblacional de los factores de riesgo que se utilizan en la función, pues sin recalibrar pueden infraestimar o sobrestimar el riesgo, incluso se recomienda recalibrar cuando se hayan modificado la incidencia de la enfermedad o la prevalencia de los factores de riesgo^{7,9,11}

Por consiguiente, los M-Ts disponibles no realizan una evaluación completa, lo que compromete una precisión predictiva del riesgo cardiovascular y no contribuyen a reducir la morbilidad y mortalidad por ECV.

El autor solo abordó, en este capítulo, las limitaciones de los M-Ts usados en la población cubana, pues en la revisión de la literatura no se encontró ningún M-IA aplicado a esta población y, por tanto, en el capítulo Resultados y Discusión se abordan algunas limitaciones de los M-IAs.

1.3 Inteligencia artificial

Inteligencia artificial, machine learning y minería de datos

La inteligencia humana no es una dimensión única, sino un espacio profusamente estructurado de capacidades diversas para procesar la información, igualmente, la Inteligencia Artificial utiliza muchas técnicas diferentes para resolver una gran variedad de tareas.^{20,86}

La literatura especializada de alto impacto, revisada en español usa indistintamente los términos *machine learning* o aprendizaje automático por lo que, el investigador también consideró su uso en esta obra escrita.

En la actualidad existe una gran cantidad de áreas específicas en las que se trabaja bajo el concepto de inteligencia artificial, que es una tecnología tan amplia y revolucionaria que resulta difícil dar una definición precisa, puede definirse como una serie de algoritmos y técnicas que pretenden imitar la inteligencia humana, sus dos técnicas principales, son el *machine learning* y el *deep learning* o aprendizaje profundo.⁸⁷

El *machine learning* comprende diferentes tipos de algoritmos, que pueden procesar gran cantidad de datos y transformarlos en conocimiento, con una intervención humana mínima. La finalidad de un algoritmo de ML es que, tomando como entrada patrones de datos, puede generar un modelo que contiene el conocimiento extraído.^{88,89}

El *deep learning*, es una técnica del *machine learning*, basado en redes neuronales artificiales con muchas capas, que son modelos matemáticos inspirados en el comportamiento biológico de las neuronas, una red neuronal artificial realiza el aprendizaje,

identificando conexiones y alterando los datos introducidos para conseguir los mejores resultados.^{15,17,87}

El aprendizaje describe la búsqueda automática de mejores y más útiles predicciones para los datos del modelo algorítmico, esto significa que la máquina puede mejorar en su tarea programada y automatizada a partir de ejemplos que se le facilitan, pero no que adquiera conocimiento o sabiduría, a pesar de lo que el término aprendizaje pueda implicar.⁸⁸

La minería de datos es un proceso de extracción de información y búsqueda de patrones de comportamiento que a simple vista se ocultan entre grandes cantidades de información:^{15,89} La definición de Minería de Datos puede variar entre los diferentes investigadores, se considera la definición de Piatetsky-Shapiro y Frawley...“La minería de datos es el conjunto de técnicas y herramientas aplicadas al proceso trivial de extraer y presentar el conocimiento implícito, previamente desconocido, potencialmente útil y humanamente comprensible, a partir de grandes conjuntos de datos, con el objetivo de predecir de forma autorizada tendencias, comportamientos y/o descubrir de forma automatizada modelos previamente desconocidos”, popularizándose esta definición por los expertos de inteligencia artificial y *machine learning*.⁹⁰

La minería de datos es la etapa de análisis del descubrimiento de conocimiento en bases de datos, (**KDD**, por sus siglas en inglés) un proceso metodológico que implica un conjunto de etapas definidas para el tratamiento de estos antes de aplicar las técnicas de minería de datos en la búsqueda de patrones ocultos en ellos y finalmente hacer el análisis de los patrones encontrados, para obtener un modelo válido, útil y entendible que los describa de acuerdo con la información y como modelo se entiende, la representación que intenta explicar ese patrón en los datos. ^{15,89,90,91}

Si el modelo final no supera la evaluación, el proceso **KDD** se podría repetir desde el principio o si el experto lo considera oportuno, a partir de cualquiera de sus etapas. Esta retroalimentación se podrá repetir cuantas veces se considere necesario hasta obtener un modelo útil y válido, por ser un proceso iterativo, ya que todas las etapas cuentan con una conexión hacia atrás, e interactivo, con la intervención del experto en el dominio de la aplicación en la toma de decisiones. En fin, la minería de datos, es considerada la etapa más importante del proceso **KDD**, todas las etapas de este proceso, serán descritas en el capítulo correspondiente.^{15,89,90,91}

La tarea de minería de datos real es el análisis automático o semiautomático de datos, donde se utilizan técnicas de la inteligencia artificial, del aprendizaje automático, la estadística y sistemas de bases de datos. Dichas técnicas, no son más que algoritmos, más o menos sofisticados que se aplican sobre un conjunto de datos para resolver un problema específico. Entre las tareas de minería de datos más importantes están la clasificación, la segmentación o clustering, la asociación y los patrones secuenciales.^{92,93}

Las técnicas de Minería de Datos se clasifican en dos grandes categorías: supervisadas o predictivas y no supervisadas o descriptivas, una técnica constituye el enfoque conceptual para extraer la información de los datos y en general es implementada por varios algoritmos.^{15,89,90,93}

En el aprendizaje inductivo supervisado existe un atributo especial, normalmente denominado clase de decisión, presente en todos los ejemplos que especifica si el ejemplo pertenece o no a la misma, es decir, se parte de un conocimiento ya dado, donde se puede predecir una salida determinada.^{15,89,90,94}

La clasificación es la técnica de ML más utilizada y permite obtener resultados a partir de un proceso de aprendizaje supervisado en el cual se encuentran propiedades comunes entre un conjunto de objetos de una base de datos y se los cataloga en diferentes clases, de acuerdo con el modelo de clasificación. ^{15,95}

Análisis predictivo

El análisis predictivo es un área de la minería de datos que consiste en la extracción de información existente en los datos y su utilización para predecir tendencias y patrones de comportamiento, pudiendo aplicarse sobre cualquier evento desconocido, se fundamenta en la identificación de relaciones entre variables en eventos pasados, para luego explotar dichas relaciones y predecir posibles resultados en futuras situaciones, y tomar así mejores decisiones. ^{96,97}

A diferencia de la estadística, en la minería de datos no se trabaja con muestras de la información, sino que se buscan patrones ocultos en los datos mediante métodos algorítmicos y utilizando un sistema automatizado. ^{15,96,97}

El ML proporciona las técnicas de análisis de datos mediante las cuales se pueden descubrir relaciones entre variables que en un principio pueden parecer insignificantes. Una vez que se han establecido correlaciones entra en juego la labor del ser humano, que consiste en saber interpretar las mismas y hacer las suposiciones apropiadas. ^{15,96,97}

El último paso, una vez encontradas e interpretadas correctamente las correlaciones, consiste en crear el modelo predictivo, se introducirán los datos de un individuo en el modelo y se obtendrá una clasificación que indicará la probabilidad de que se produzca la situación estudiada. ^{15,89,91,95}

Cuando se habla de análisis predictivo realmente se refiere a la creación de modelos predictivos, no obstante, el auge que la analítica predictiva está experimentando ha hecho que el término se utilice para referirse a todo lo relacionado con la disciplina del análisis de datos.

Las técnicas de IA, son tecnologías de gran potencialidad en la medicina, ante la gran cantidad de datos que no puede ser procesada y analizada con métodos tradicionales con el fin de generar correlaciones, tendencias y patrones, los complejos modelos algorítmicos de ML han demostrado ser más precisos que los tradicionales como una nueva forma de generar conocimiento.^{15,16,19,20}

1.4 Conclusiones del capítulo 1

Las enfermedades cardiovasculares representan una carga económica, social y humana en la mayoría de los países del mundo, incluida Cuba. La aterosclerosis, y su fase subclínica prolongada, es el proceso subyacente de la enfermedad cardiovascular, que junto a la dislipemia aterogénica y el síndrome metabólico, tienen en común la inflamación de bajo grado y, a pesar de sus criterios diagnósticos establecidos, son infradiagnosticadas e infracontroladas en la atención primaria de salud. La predicción del riesgo cardiovascular, es una herramienta imprescindible en la práctica médica, no obstante, se necesita añadir factores de riesgo emergentes en los modelos predictivos para una evaluación completa del riesgo cardiovascular. El análisis histórico-lógico, permitió revelar insuficiencias teórico-metodológicas, en los M-Ts disponibles, por tanto, esos modelos no estiman eficazmente la probabilidad de tener un episodio cardiovascular y, por consiguiente, no contribuyen a la reducción de la morbilidad y mortalidad por enfermedades cardiovasculares. Por último, las técnicas de la inteligencia artificial se avizoran como un campo promisorio en medicina y

salud, debido a que podrían contribuir a la predicción, prevención y tratamiento de las enfermedades, como las cardiovasculares en la atención primaria de salud.

Las consideraciones expuestas en este capítulo constituyen la base teórica para la correcta valoración del problema que se investiga, así como la adecuada interpretación de sus resultados.

CAPÍTULO 2. DISEÑO METODOLÓGICO

En este capítulo se presenta la estrategia metodológica para dar cumplimiento a los objetivos del estudio: se describen las tres etapas estrechamente relacionadas de la investigación: el ensamble y seguimiento del estudio de cohorte prospectivo, que fue la fuente de los datos; el diseño y validación del modelo predictivo, usando la metodología KDD, empleando la herramienta de minería de datos Weka y finalmente, en la tercera etapa se comparó el modelo obtenido con técnicas de inteligencia artificial con las tablas predictivas de la OMS de 2019 para las Américas, región del Caribe; además se muestran, elementos metodológicos como la clasificación del estudio, el universo y la muestra, los criterios de inclusión, exclusión y salida, el ensamble y seguimiento de la cohorte, las etapas del proceso de Descubrimiento de Conocimiento en Base de Datos que son selección de datos, preprocesamiento, transformación, minería de datos e interpretación, evaluación y validación de los modelos predictivos; se hace una descripción de los métodos de clasificación, algoritmos de *machine learning* y las técnicas de evaluación de los atributos y modelos, la operacionalización de los atributos predictores y el atributo o clase de decisión, así como, los métodos empleados en la investigación, el análisis estadístico y los aspectos éticos.

2.1 Diseño general de la investigación

2.1.1 Estudio de cohorte prospectivo PredRCG_aps

Clasificación de la investigación

En la primera etapa fue realizado un estudio observacional, analítico de cohorte prospectivo, denominado por el autor PredRCG_aps⁶³ en pacientes atendidos en el Policlínico “José Ávila Serrano”, en la localidad de Velasco, perteneciente al municipio de Gibara, en la

provincia de Holguín, en Cuba, la cohorte fue seguida desde enero de 2008 hasta diciembre de 2019

Características del universo y muestra

El policlínico docente “José Ávila Serrano” atendía, en el año 2008, un universo poblacional en el rango de 35 a 74 años, de 16 054 habitantes aproximadamente, 7947 del sexo masculino y 8107 del femenino.

Determinación del marco muestral

En una primera etapa se pesquisó la población de 35 a 74 años sin historia de enfermedad cardiovascular establecida como: infarto del miocardio o angina, enfermedad cerebrovascular o isquemia cerebral transitoria, enfermedad arterial periférica, insuficiencia cardíaca o arritmias.

Determinación del tamaño muestral

Se contrastó el efecto del factor por medio de un test de Wald en un modelo de regresión de Cox ajustado por varios factores, usando el paquete estadístico Epidat 4.2 con los siguientes parámetros: riesgo relativo a detectar (2); proporción de expuestos (0,5); proporción observaciones censuradas (0,8); nivel de confianza (0,95); poder estadístico (0,8); correlación con los otros factores a incluir en el modelo (0,28); se determinó un tamaño muestral mínimo de 355 pacientes.⁶³

Criterios de inclusión y exclusión

Se incluyeron: Pacientes de ambos sexos con edades entre 35 y 74 años, no portadores de una enfermedad cardiovascular establecida como: infarto del miocardio o angina;

enfermedad cerebrovascular o isquemia cerebral transitoria; enfermedad arterial periférica, insuficiencia cardíaca o arritmias.

Se excluyeron: Pacientes portadores de otras enfermedades crónicas invalidantes, trastornos psiquiátricos; enfermedades inflamatorias, autoinmunes, vasculitis o cáncer, donde habitualmente se observan valores elevados de la proteína c reactiva de alta sensibilidad mayores de 10 mg/dL.

Criterios de salida: voluntad expresa del paciente de no continuar en la investigación, muerte por causa no cardiovascular, abandono del país o pérdida en el seguimiento.

Por ser un estudio con casos censurados, es decir, el paciente no desarrolló el evento cardiovascular en el periodo de seguimiento, que fue la variable independiente; en caso de abandono, muerte o pérdida del seguimiento por otra causa, fue utilizada toda la información disponible de esos pacientes.

Los criterios de inclusión y exclusión garantizaron que los participantes fueran representativos de la población objetivo prevista.

Ensamble y seguimiento de la cohorte

Ensamble de la cohorte

La cohorte prospectiva PredRCG_aps, fue ensamblada durante los años 2008 y 2009

La población de 35 a 74 años sin historia de enfermedad cardiovascular establecida, fue seleccionada mediante un muestreo probabilístico aleatorio simple, lo que aseguró también, que la cohorte fuera representativa de la población objetivo estudiada, los pacientes incluidos fueron sometidos a un interrogatorio, examen físico, antropométrico y a

determinaciones de laboratorio previo un ayuno de 12 a 14 horas, para la medición de las variables basales.

La cohorte ensamblada, con una fracción de esa población, incluyó 857 pacientes del sexo femenino y 776 del masculino, la información fue recogida a la inclusión y en el seguimiento anual del paciente (Anexo 2) y almacenada en una base de datos en formato Excel, con 16 variables, cuatro nominales y el resto continuas, evaluadas por su asociación con el resultado de interés, como : edad (edad), sexo, consumo de tabaco (hfumar), antecedentes personales de diabetes mellitus (appdm) e hipertensión arterial (apphta), Índice de Masa Corporal (IMC), circunferencia abdominal (circabd), tensión arterial sistólica (tas) y diastólica (tad), colesterol total (coltot), colesterol unido a lipoproteínas de baja densidad (cLDL), colesterol unido a proteínas de alta densidad (cHDL), triglicéridos (triglic o TG) también denominado triacilglicéridos en la literatura especializada, glucemia o glicemia en ayunas (glicay), ácido úrico (aciduric) y proteína c reactiva de alta sensibilidad (pcr-as o pcr-hs), determinada por método inmuniturbidimétrico, este método, unido a su larga vida media y su escasa variación en tiempo, hacen que sea el marcador de inflamación más ampliamente utilizado y la variable resultado que fue el desarrollo de un evento cardiovascular mortal o no.

En el anexo 3 se describe, del estudio de cohorte prospectivo, la operacionalización de las variables independientes o también nombradas predictores candidatos, que son aquellos considerados potencialmente predictores de la ocurrencia del resultado y la variable resultado, como ha sido mencionado fue el desarrollo de un evento cardiovascular mortal o no.

Los exámenes de laboratorios fueron realizados por el mismo personal especializado de la institución, de Inmunología de la Universidad de Ciencias Médicas de Holguín y por el laboratorio central del Hospital Provincial Docente “VI Lenin”, de Holguín, Cuba.

Seguimiento de la cohorte

La cohorte fue seguida por un periodo de $10,6 \pm 1,3$ años hasta diciembre de 2019, se realizaron evaluaciones clínicas anuales de cada paciente. (Anexo 2) Fue verificado que el efecto de la variable se mantenía constante a lo largo del tiempo, evaluándose visualmente con un gráfico log-log en la variable numérica y ordinal, fue utilizada la última información disponible en la evaluación del paciente.

El resultado fue la ocurrencia de un evento cardiovascular (variable respuesta, resultado o dependiente), como angina de pecho, infarto del miocardio, enfermedad cerebrovascular isquémica o hemorrágica, crisis de isquemia cerebral transitoria y enfermedad arterial periférica, estos resultados se sustentan en las guías clínicas como “eventos duros” que no involucran el juicio subjetivo; se realizó la revisión del registro de su médico de familia, de la historia clínica hospitalaria y en caso de evento mortal, el de mortalidad municipal. Se empleó la Clasificación Internacional de Enfermedades (CIE- 10), se definieron las causas de muertes como enfermedades cardíacas (CIE-10: 120, 121-122, 124), enfermedades cerebrovasculares (CIE-10: 160-162,163,164) y otras de tipo cardiovasculares (CIE-10: 173).⁶³

El estudio de cohorte fue conducido por el mismo investigador, excepto dos años por otra investigadora. Como la evaluación fue anual los datos faltantes fueron mínimos: 82 pacientes. El autor utilizó el método de imputación múltiple, que consistió en estimar los

valores ausentes sobre la base de los válidos de otras variables y/o casos de la muestra, evitando un rendimiento sesgado por valores faltantes.

La investigación fue realizada en el contexto de la práctica médica habitual, durante el periodo de seguimiento, los pacientes no fueron privados de los protocolos establecidos en la atención primaria de salud en Cuba.

2.1.2 Diseño y validación del modelo predictivo. Metodología KDD

La segunda etapa de la investigación, fue retrospectiva porque partiendo de las causas se conoce el efecto, usando técnicas de minería de datos supervisadas o predictivas, lo que permitió la extracción de conocimiento con el proceso conocido como KDD^{12,89,91} por sus siglas en inglés, que significa descubrimiento o extracción de conocimientos en bases de datos, esta metodología (Anexo1, figura 2) permitió llevar a cabo el proceso de diseño de los modelos predictivos de forma sistemática y se dividió en 5 etapas: selección de datos, preprocesamiento, transformación. minería de datos, evaluación y validación.

Fue utilizada la herramienta de minería de datos WEKA, versión 3.9.5,⁹⁸ herramienta de código abierto escrita en JAVA, disponible bajo licencia publica GNU, que implementa *machine learning* y que tiene como objetivo brindar algoritmos de minería de datos y permite la obtención de conocimiento mediante el proceso KDD. (Anexo 1, Figura 3) A continuación, se describen las etapas del proceso KDD.

2.1.2.1 Selección y carga del conjunto de datos en Weka

Selección del conjunto de datos: La fuente de datos fue la base de datos (base de casos), del estudio de cohorte prospectivo PredRCG_aps, con 1633 pacientes y 16 variables, todas con una relación monótona con el evento de interés.

Carga del conjunto de datos: La base de datos en formato Excel para cargarla a Weka fue codificada en formato cvs y el fichero creado se nombró riesgocardv1.arff, con cada instancia en una fila y con los atributos separados por comas.

Construcción del archivo arff

El *software* trabaja con un formato de archivo que se denomina arff (***A*tribute *R*elation *F*ile *F*ormat**, por sus siglas en ingles), se muestra una fracción del archivo en la figura 4 (Anexo 1) y que consta de tres partes:

1. Cabecera: donde se define el nombre del archivo a través de la expresión *@relation* <nombre-de-la-relación>.
2. Declaración de atributos: se corresponde con el segundo bloque y es donde se definen los atributos que vamos a estudiar, así como su tipo, por medio de la expresión *@attribute*<nombre-del-atributo><tipo>.
3. Sección de datos: es la parte final de archivo, se inicia con *@data* y es el lugar donde se insertaron los datos que componen nuestra base de datos, los atributos deben estar separados por comas y con saltos de línea cada relación (instancias).

Delimitación y operacionalización de los atributos o rasgos predictores y el rasgo o clase de decisión

La base de datos, en formato arff denominada *@relation riesgocardv1*, quedó conformada por 1633 instancias y 16 atributos o rasgos predictores, 4 nominales o categóricos y el resto numéricos (Anexo 3, tabla 2) y un atributo especial o rasgo de decisión.

Atributo o rasgo de decisión

Fue denominado @evento, cuyo dominio fue {Evento= Sí, Evento= NO}, clase binaria, donde se indica si el paciente desarrolló o no el evento cardiovascular. La base de datos tenía dos clases de decisión: 1531 pacientes no desarrollaron el evento cardiovascular y 102 pacientes sí, de las dos clases de decisión, los que desarrollaron el evento está menos representada, lo que indica que los datos tenían un alto desbalance.

2.1.2.2 Preprocesamiento

Esta etapa consistió en la preparación y limpieza de los datos con valores redundantes, incorrectos o faltantes, Weka también dispone de un filtro no supervisado dirigido a instancias nombrado **ReplaceMissingValues**, que permitió reemplazar valores perdidos, se validaron los datos, para un manejo estandarizado, que permitió obtener mejores resultados según las características del algoritmo de clasificación a utilizar, aplicándose filtros supervisados sobre los atributos o las instancias.

Los datos originales presentaban un alto desbalance, con una razón (*Imbalanced Ratio*, IR en inglés) de 15. En la figura 5 (Anexo 1) se aprecia en azul la clase Evento =NO y en rojo la clase Evento=Sí, que es la clase de interés, que permitió balancearlas, buscando proporción entre los pacientes que desarrollaron el evento cardiovascular versus los que no; se aplicaron técnicas de remuestreo, con el algoritmo supervisado dirigido a las instancias, **SMOTE** (*Synthetic Minority Over-sampling TEchnique*), aumentándose la clase minoritaria con una técnica de sobremuestreo (*over sampling*), por cada ejemplo de la clase minoritaria, se introdujeron ejemplos sintéticos en la línea que une al elemento con sus cinco vecinos más cercanos.⁹⁹

2.1.2.3 Transformación

En la transformación se convirtieron, normalizaron o discretizaron los datos, para un manejo homogéneo y que el proceso de minería pudiera ser más eficiente y, permitiera obtener mejores resultados o para poder utilizar algoritmos que requieren ciertas características en el formato de los atributos predictores.

Para el análisis predictivo, disponiendo de atributos numéricos se conformó otro archivo. arff, donde los atributos predictores fueron codificados como simbólicos en la herramienta Weka, que serán descritos en la sección de Resultados, pues algunos algoritmos predictores trabajan con un formato específico de atributos.

2.1.2.4 Minería de datos

Selección de atributos significativos

Se emplearon técnicas de selección de atributos integradas en Weka para identificar los atributos más significativos, que se utilizaron en la generación de los modelos predictivos como: **CfsSubsetEval**, con el método de búsqueda **BestFirst**, **CorrelationAttributeEval**, con el método **Ranker** y **WrapperSubsetEval/J48/BestFirst**, utilizando la validación cruzada con 10 particiones (*10 folds cross-validation*), se consideraron significativos los atributos que alcanzaron un valor relevante en dos o más métodos de selección. También es necesaria la experiencia de expertos en el área para la selección de atributos, que en la investigación fue el autor, para de esta forma, no eliminar atributos que pudieran ser relevantes en las predicciones.

Aprendizaje supervisado. Técnicas predictivas

Fueron utilizados algoritmos clásicos, de alto desempeño predictivo, propios de Weka que implementan las funcionalidades necesarias para generar los modelos predictivos.

Se ejecutaron algoritmos *machine learning* de clasificación y fueron configurados los hiperparámetros, antes del proceso de entrenamiento del modelo *machine learning* para optimizar su rendimiento y mejorar su capacidad de generalización.

Los algoritmos ejecutados fueron: Algoritmo de clasificación de reglas: **JRip**, algoritmo de clasificación de árboles: **J48**, algoritmo de clasificación de funciones: **Multilayer Perceptron with Backpropagation** y el meta-algoritmo de clasificación: **AttributeSelectedClassifier**, con el método de búsqueda **BestFirst** y el evaluador **CfsSubsetEval**. Los algoritmos de clasificación se ejecutaron a los datos originales y en los datos balanceados.

2.1.2.5 Evaluación y validación. Medidas de rendimiento del clasificador

En esta etapa se realizó la interpretación, evaluación y validación de los modelos obtenidos, se identificaron y compararon estos y, además, se procedió a su validación comprobando que las conclusiones que arrojan son válidas y los modelos generalizan bien. En la investigación se consideró un problema de clasificación binaria, dada la clase de decisión @evento, cuyo dominio es {Evento= Sí, Evento= NO}, la clase minoritaria Evento= Sí, fue considerada como la clase positiva y la clase mayoritaria Evento= NO, como la clase negativa.

Recordemos que, la matriz de confusión para dos clases que se muestra en la figura 6 (Anexo 1); las columnas de esta matriz están asociadas al número de predicciones de cada clase asignadas por el clasificador y las filas al total de instancias de la clase real, TP (*True Positive*, verdaderos positivos) y TN (*True Negative*, verdaderos negativos), que constituyen el número de objetos correctamente clasificados en las clases Evento= Sí (clase positiva) y Evento=No (clase negativa) y por otro lado, FP (*False Positive*, falsos

positivos) y FN (*False Negative*, falsos negativos) , son el número de objetos mal clasificados de las clases Evento= SÍ y Evento = No, respectivamente.

Se aplicaron las métricas basadas en la matriz de confusión, cuantificando la calidad del algoritmo clasificador empleado en el aprendizaje supervisado inductivo, permitiendo, la comparación y selección de los modelos creados.

Para la evaluación del rendimiento del clasificador fueron utilizadas las métricas más usadas para clases desbalanceadas, como: la tasa de verdaderos positivos (*TPrate*, en inglés) y tasa de falsos positivos (*FPrate*, en inglés), en la clase de interés, que fue la minoritaria, es decir Evento= SÍ, el área bajo la Curva de Operación del Receptor (ROC), la Curva Precisión _ Recall (PRC), la Medida F (*F Measure*, en inglés) y el Coeficiente de Correlación de Matthew (MCC). (Anexo 4)

Validación de los modelos predictivos

El *software* WEKA cuenta con cuatro opciones para el proceso de validación de modelos, al disponer de una sola base de datos en la investigación, se utilizó el conjunto de datos para entrenamiento y para la validación se utilizó la técnica de validación cruzada con 10 particiones (*k-fold cross validation*): se dividió aleatoriamente el conjunto de datos en 10 partes (folds) y se realizaron 10 iteraciones en las que se fue reservando aleatoriamente, una parte como datos de prueba y el resto como datos de entrenamiento y al final se promediaron los resultados de las precisiones y errores calculados en todas las iteraciones, se muestra una fracción del proceso. (Anexo 1, figura 7)

Los modelos predictivos fueron sometidos a una evaluación crítica por el autor e independiente por los tutores, fue aplicada la herramienta PROBAST (*A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies*), herramienta para evaluar el riesgo de

sesgo y la aplicabilidad de los modelos predictivos, que consta de 4 dominios que contienen 20 preguntas de señalización para facilitar la evaluación del RoB (*Risk of Bias, Riesgo de Sesgo*) y la aplicabilidad de los modelos.¹⁰⁰

Las preguntas de señalización de PROBAST, fueron extraídas utilizando la herramienta CHARMS (*Checklist for critical Appraisal and data extraction for automated Reviews of prediction Modelling Studies*)¹⁰¹, además, fue aplicada la lista de verificación (*checklist*) de las guías TRIPOD-AI (*Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis*), adaptadas a la inteligencia artificial, que permitió mejorar la transparencia en los informes para facilitar la reproducción de modelos y aumentar la comprensión y la aplicabilidad de los M-IAs para una práctica clínica más amplia.^{3,102}

2.1.3 Comparación del modelo predictivo implementado con las Tablas de Predicción del Riesgo Cardiovascular de la OMS de 2019

En la tercera etapa de la investigación se comparó el modelo predictivo implementado, con las tablas de riesgo de la OMS, recalibradas en el 2019, para 21 regiones epidemiológicas, las tablas de la región del Caribe, que son el modelo válido para su uso en la población cubana según recomendaciones del paquete técnico HEARTS.¹⁴

Fue implementado el M-IA nombrado por el autor como “ModPred IA EnfCardiovasc_aps 1.0”, utilizando los modelos predictivos generados por los algoritmos de clasificación de reglas **JRip** y el de clasificación de árboles **J 48**, que generan reglas de decisión que siguen una sintaxis condicional del tipo: **SÍ-ENTONCES (IF-THEN)** , **SÍ** se cumplen las condiciones (también llamados antecedentes), **ENTONCES** se hace la predicción, en la investigación si

se cumple la regla de decisión el paciente se considera en riesgo de ECV y de no cumplirse ninguna, no se clasifica como riesgo.

La tabla de predicción de riesgo de la OMS¹ utiliza como predictores: la edad, el consumo de tabaco, la diabetes mellitus, las cifras de tensión arterial sistólica y de colesterol total y, otro modelo con datos no analíticos cuando no se dispone de colesterol y glucemia, donde se incluye el IMC y no se usa la Diabetes Mellitus, también, se reduce el nivel de riesgo en comparación con las tablas de predicción OMS del año 2007.⁷⁷

Fue comparado el M-IA implementado, con las tablas de la OMS de 2019, según: tipo de estudio, método empleado, nivel de evidencia de estudios pronósticos, validación, categorías de riesgo, edad de los participantes, número y tipo de variables, aplicabilidad y facilidades de uso.

Fue comparada la predicción de ambos modelos usando la base de datos original, sin balancear, de la cohorte prospectiva, también, fue aplicado el índice aterogénico TG/cHDL, con un punto de corte mayor de 2 que indica mayor número de partículas LDL oxidadas, pequeñas y densas, que son las más aterogénicas e influyen en la formación de placas de ateromas, que originan la aterosclerosis y las ECV, además, se determinaron las concentraciones de pcr-as mayores de 1mg/dL en los pacientes que las tablas de la OMS de 2019 clasificaron como riesgo bajo e intermedio y el M-IA como riesgo Sí.

Todos los experimentos fueron ejecutados en una computadora ASUSTek, modelo X543SA, con las siguientes características técnicas: procesador Intel® Celeron R @ 1,60GHz 1,60 GHz, 4 GB en RAM, sistema operativo Windows 10 de 64 bits, procesador x64.

2.2 Métodos empleados en la investigación

Del nivel teórico: Interpreta los datos empíricos, crean condiciones para ver las características de la realidad, explican los hechos y profundizan en las relaciones esenciales y cualidades de los procesos observables o no directamente.

- ✓ Análisis documental o de literatura: Posibilita identificar los referentes sobre el tema de investigación, que sustentan los planteamientos hechos por el autor.
- ✓ Histórico-Lógico: Lo histórico estudia la trayectoria de los fenómenos y acontecimientos históricos, y lo lógico, investiga las leyes generales del funcionamiento y desarrollo de los fenómenos.
- ✓ Análisis y Síntesis: En el análisis se descompone el objeto en sus distintas partes, revelando sus componentes y relaciones; en la síntesis, se establece mentalmente la unión de las partes previamente analizadas y posibilita descubrir relaciones y características generales entre sus componentes.
- ✓ Inducción y Deducción: La inducción es una forma de razonamiento a través del cual se pasa de un conocimiento de cosas particulares a un conocimiento más general que va a reflejar lo que hay de común en esos fenómenos individuales, y en la deducción se pasa de un conocimiento general a un conocimiento particular y/o conocimiento menos general.
- ✓ Hipotético Deductivo: Permite adelantar y verificar las nuevas hipótesis sobre la realidad, establece nuevas predicciones a partir del sistema de conocimientos que se tiene.
- ✓ De lo abstracto a lo concreto: Lo abstracto destaca la relación de las cosas y fenómenos, y lo concreto, es la síntesis de muchos conceptos y, por consiguiente,

de las partes. Mediante este método se posibilita aislar, separar y determinar las cualidades esenciales de cada objeto.

- ✓ Modelación: Se reproduce la realidad objetiva que se está estudiando como objeto, se descubren y estudian nuevas cualidades, relaciones, principios o leyes del objeto de estudio a partir del proceso de abstracción que se desarrolla.
- ✓ Método dialéctico: Se investiga la verdad a través del examen crítico de las percepciones y teorías mediante el intercambio de proposiciones y contraproposiciones, resolviendo la contradicción a través de la formulación de una conclusión final.

Del nivel empírico: Posibilitan estudiar fenómenos, objetos, y procesos observables y poderlos confirmar mediante hipótesis y las teorías.

- ✓ Medición: dirigido a precisar información numérica acerca de la cualidad del objeto mediante comparaciones de magnitudes determinadas.
- ✓ Encuesta: Permite caracterizar en la población que se estudia las variables de interés. Se aplicó por el autor un cuestionario al inicio de la investigación y en cada año de seguimiento. (Anexo 2)

2.3 Análisis estadístico

Para la determinación del tamaño muestral mínimo para el ensamble de la cohorte fue utilizado el paquete estadístico Epidat 4.2

Fueron utilizadas media aritmética \pm desviación estándar si las variables eran de tipo continuo y de frecuencia relativa cuando eran categóricas, así como las redes neuronales

con la función *multilayer perceptron* con el software *IBM Statistical Package for Social Sciences* (IBM SPSS), versión 25 para *Windows*.

En la minería de datos se utilizaron técnicas de la inteligencia artificial, algoritmos de *machine learning* y técnicas estadísticas de modelización que permitieron extraer información del conjunto de casos y transformarla en una estructura comprensible para su uso posterior.

2.4 Ética

La dirección del centro y el Comité de Ética de la Investigación Clínica dieron su aprobación para la ejecución de la investigación. (Acuerdo 16/2008).

La investigación desde el punto de vista ético cumplió con la Declaración de Helsinki para las investigaciones médicas en seres humanos.¹⁰³

Para mantener la confidencialidad, los datos de la investigación fueron manejados exclusivamente por el responsable del proyecto. Para la inclusión de un paciente en el estudio se obtuvo el consentimiento informado, firmado por el paciente y el investigador. (Anexo 5) Se explicó a cada participante toda la información relevante sobre los objetivos y métodos de la investigación, se preservó la privacidad de la información y se protegió la base de datos del acceso de otras personas ajenas a la investigación, se permitió la libertad de retirarse del estudio si lo deseaban. Como la investigación fue realizada en el contexto de la práctica médica habitual, a ningún paciente, durante el periodo de seguimiento, se privó de los protocolos establecidos en la atención primaria de salud.

Consideraciones finales del capítulo 2

En este capítulo se presentaron las tres etapas del estudio: la fuente de datos, que fue el estudio de cohorte prospectivo; la segunda etapa, el diseño del modelo predictivo, usando la metodología KDD o Descubrimiento de Conocimiento en Bases de Datos, se describieron las etapas de la metodología KDD. Además, se identificaron los algoritmos ejecutados en el proceso de descubrimiento de conocimiento, la técnica de preprocesamiento aplicada en clases desbalanceadas, la selección de atributos significativos, la transformación de numéricos a categóricos. Se precisaron algunas medidas para evaluar los clasificadores en clases desbalanceadas que permitieron reflejar la calidad de la clasificación, considerándose principalmente el área bajo la curva ROC, el área bajo la curva Precisión_Recall (PRC), la Medida F (*F measure*) y el Coeficiente de Correlación de Matthew (MCC) y fue realizada la validación cruzada con 10 particiones. La metodología de extracción del conocimiento en bases de datos KDD y la herramienta de minería de datos Weka, están entre las más populares y, por consiguiente, más usadas y dieron cumplimiento al primer objetivo de la investigación. Se aplicaron instrumentos para determinar el riesgo de sesgo, la aplicabilidad y la transparencia de los modelos predictivos. Finalmente, en la última etapa, se procedió a comparar el M-IA implementado con las tablas de la OMS, recalibradas en 2019, recomendadas por el paquete técnico HEARTS para su uso en la población cubana.

CAPÍTULO 3. ANÁLISIS DE LOS RESULTADOS Y DISCUSIÓN

Este capítulo expone los resultados de la investigación y su discusión, cuya estructura se corresponde con los objetivos y las etapas del diseño de la misma.

3.1 Resultados y Discusión. (Anexo 6)

3.1.1 Estudio de cohorte prospectivo PredRCG_aps

La cohorte prospectiva, que incluyó 1633 personas, 776 masculinos y 857 (52,5 %) femeninos, nos permitió determinar la incidencia de los eventos cardiovasculares en el periodo de seguimiento, así como la prevalencia de los factores de riesgo cardiovasculares. Durante el seguimiento en el sexo masculino, 59 pacientes presentaron un evento cardiovascular (7,60 %) y 43 en el femenino (5 %), y el 6,2 % en la cohorte.

Este estudio se basó en una fracción de la población, con un periodo de seguimiento de \pm 10 años, que se consideró suficiente para que todos los sujetos de la cohorte tuvieran la opción de desarrollar o no el evento de interés. El estudio fue prospectivo, el registro de los datos “hacia adelante” lo hace que sea más efectivo; es decir los sujetos, podían estar expuestos o no al factor de riesgo, pero el evento de interés, aún no había ocurrido al ensamblar la cohorte.⁶³

Los estudios de cohorte constituyen un valioso recurso para la prevención y el tratamiento de las enfermedades y los puntajes de riesgo, basados en cohortes poblacionales, son herramientas muy útiles en la práctica clínica, ya que permiten clasificar a las personas según grupos de riesgo y, por tanto, priorizar las intervenciones de prevención en sujetos con un riesgo cardiovascular más elevado.^{8,63}

La prevalencia de factores de riesgo cardiovascular seleccionados, algunos no incluidos en las tablas de predicción actuales (Anexo 6, gráfico 3) según sexos y en la cohorte, muestra una elevada prevalencia del antecedente de hipertensión arterial (27,7 %), del nivel de proteína c reactiva de alta sensibilidad igual o mayor de 3 mg/dL (21,5 %), lo que es indicador de una inflamación crónica o de bajo grado,^{104, 105} del consumo de tabaco (16,9 %), la obesidad (16,7 %) y la obesidad abdominal (9,6 %), así como las cifras de glucemia alterada en ayunas en personas no diabéticas (10 %) y triglicéridos mayor de 2,2 mmol/L (17,2 %). El autor considera deficiente el control de estos factores de riesgo cardiovascular en la atención primaria de salud, en un sistema de salud como el cubano enfocado en el cuidado primario, donde la promoción y prevención de enfermedades mediante la detección de factores de riesgo debía ser la prioridad, con la consiguiente modificación del estilo de vida y otras estrategias oportunas para su control.

En estudio de cohorte prospectivo realizado en Cuba, ensamblado de 1996 a 2002, en cinco provincias, que reclutó a 146 556 adultos con edad mayor o igual de 30 años y seguidos hasta 2017 mediante el monitoreo electrónico del registro de mortalidad del Ministerio de Salud Pública, la mortalidad cardiovascular se asoció significativamente con la tensión arterial sistólica ≥ 120 mm Hg, la diabetes mellitus tipo 2 y un índice de masa corporal ≥ 22.5 kg/m².¹⁰⁶ La prevalencia de los factores de riesgo en la población cubana, según la encuesta nacional de salud, coincide con lo reportado en el estudio de cohorte prospectivo.¹⁰⁷

3.1.2 Diseño y validación del modelo predictivo. Metodología KDD

El diseño de un sistema de analítica predictiva requiere de un proceso estructurado y disciplinado, para encontrar un modelo útil, válido y entendible que describa los patrones

de acuerdo con la información. Como ha sido mencionado, el **Descubrimiento de Conocimiento en Base de Datos** (*Knowledge Discovery in Databases*, **KDD** en inglés), es ese proceso metodológico, secuencial y se entiende, como proceso un conjunto de etapas sucesivas, típico de la minería de datos, que es el núcleo del proceso KDD.⁹⁰

En la revisión de la literatura, algunos autores describen nueve etapas en el proceso KDD, la primera es la abstracción del escenario, es imprescindible un experto en el dominio de la aplicación y definir qué problema se quiere resolver y cuáles son los objetivos que se pretenden alcanzar, la selección de los datos, su limpieza y preprocesamiento, la transformación de los datos, la selección de la tarea de minería de datos, la elección y aplicación del o los algoritmos clasificadores, la evaluación de los modelos, su validación y, por último, la aplicación del conocimiento para resolver el problema.^{15,90,91}

La relación de todas estas etapas solo es lineal en el papel, siendo mucho más compleja y esconde una jerarquía de subfases, si el modelo final no supera la evaluación el proceso se podría repetir desde el principio o si el experto lo considera oportuno, a partir de cualquiera de los pasos anteriores. Esta retroalimentación (*feedback*) se podrá repetir cuantas veces se considere necesario hasta obtener un modelo útil y válido.^{89,90}

Etapas 1. Selección de datos

La fuente de los datos, con la que se construyeron los modelos predictivos, como se ha mencionado, fue aportada por el estudio de cohorte predictivo PredRCG_aps.⁶³

La recopilación de los datos es la piedra angular de un proyecto de minería de datos, su calidad afecta directamente la capacidad de extraer información significativa y confiable, esta fase es crítica ya que sienta las bases para las capacidades analíticas del proyecto.¹⁵

Etapa 2. Preprocesamiento

Los datos del atributo o clase de decisión tenían un alto desbalance (Anexo 1, figura 5), 102 personas desarrollaron un evento cardiovascular y 1531 no lo presentaron, para una razón de desbalance (*Imbalanced Ratio*, IR en inglés) de un 15. (1531/102), lo que produce un desbalanceo binario en este conjunto de datos formados por dos clases, ya que la mayoritaria representa los pacientes que no presentaron un evento cardiovascular y tiene un mayor número de datos con respecto a la clase de interés que es la minoritaria conformada por los pacientes que sufrieron un evento cardiovascular.¹⁰⁸

El problema de las distribuciones de datos no balanceados entre clases, ha recibido una atención considerable en disciplinas como el *Machine Learning* y la Minería de Datos. El principal problema que ocasiona una base de datos desbalanceada, es que los algoritmos clasificadores tienden a sesgar sus resultados en favor de la clase mayoritaria, ya que estos algoritmos están contruidos para operar en datos balanceados.^{15,106,109,110}

Para balancear los datos se aplicaron técnicas de remuestreo (*oversampling*), fue aplicado un filtro supervisado dirigido a instancias con el algoritmo **SMOTE**, se introdujeron instancias sintéticas en la clase minoritaria para equilibrarla con la mayoritaria, así las nuevas instancias sintéticas se generaron a través de la interpolación de varias instancias basándose en la regla del vecino más cercano, la nueva instancia sintética se coloca entre la instancia seleccionada originalmente y sus k vecinos más cercanos,^{101, 108} quedando la clase minoritaria con 765 instancias, con un IR de 2, es decir balanceadas las clases. (Anexo 6, figura 8)

En la actualidad, el problema del desbalance es complejo, no hay una técnica mejor, todas tienen sus ventajas y limitaciones. En la investigación fueron utilizadas varias técnicas de

submuestreo, sobremuestreo e híbridas, en las técnicas *oversampling* no hay pérdida de información, aunque puede haber repetición de muestras ruidosas y mayor tiempo de pre procesado. La literatura refiere, de forma experimental, que el IR aceptable para considerar las clases balanceadas, debe ser inferior a 9, sin brindar explicaciones teóricas, unas de las limitaciones del algoritmo **SMOTE** es que puede sobreestimar en la clase minoritaria,^{113,114} pero los datos no son complejos y no se evidenció desbalance al interior de las clases^{63,113} y este algoritmo ha sido usado en muchas aplicaciones, tiene una gran popularidad y ha mostrado un alto desempeño competitivo.

Etapa 3. Transformación

Las técnicas de preparación de datos se orientan a la adecuación del juego de datos para que pueda ser usado, posteriormente por los algoritmos de clasificación.^{15,96,97}

Fue utilizado el archivo *riesgocardv1.arff* con 4 atributos nominales y el resto numéricos y se aplicó el filtro supervisado *discretize*, además se transformaron los atributos numéricos, excepto la edad, en atributos nominales o categóricos, codificándolos en Weka (Anexo 6, tabla 3), según los valores normales o deseables, en riesgo moderado o alto, considerados en la práctica médica y el archivo fue denominado *riesgocardvdummy.arff*.

La discretización es una tarea habitual en los procesos de minería de datos puesto que muchos algoritmos solo aceptan valores nominales, porque trabajar con atributos nominales en lugar de continuos, disminuye el coste computacional y acelera el proceso inductivo.^{15,97}

Etapa 4. Minería de datos. Diseño del modelo predictivo

Selección de atributos significativos

Fueron aplicados métodos de selección de atributos, incluidos en Weka, al conjunto inicial de datos sin balancear, para determinar los atributos con mayor poder predictivo. Los resultados de los métodos (Anexo 6, tabla 4) no coincidieron totalmente, ya que cada uno captura de manera diferente la información y la relevancia de las variables.

El método **CfsSubsetEval** evalúa el valor de un subconjunto de atributos al considerar la capacidad predictiva individual de cada característica junto con el grado de redundancia. Los atributos con información redundante son penalizados puesto que tendrán una alta correlación con otras características, y los irrelevantes son ignorados, puesto que tendrán una baja correlación con la clase.^{15,114}

El método **WrapperSubsetEval** evalúa conjuntos de atributos utilizando un esquema de aprendizaje. Se trata de un método de envoltura en el cual, para evaluar y seleccionar los atributos, se necesita un método de inducción, utiliza un algoritmo de clasificación para poder hacer la medición del rendimiento de los atributos que se están evaluando.

^{15,114} **CorrelationAttributeEval** evalúa el valor de un atributo midiendo la correlación (Pearson) entre él y la clase. En los atributos nominales se consideran valor por valor al tratar cada uno como un indicador. Se llega a una correlación general para un atributo nominal a través de un promedio ponderado.^{15,114,115}

Los atributos seleccionados, que alcanzaron puntuaciones medias o altas en dos o más métodos de selección, fueron los siguientes: edad, hfumar, apphta, tas, circabd, triglic, cHDL, cLDL, glicay y pcr-as, aunque Weka permite remover atributos, generalmente se eliminan los menos significativos o con menor poder predictivo; de igual forma, si se remueven los atributos con mayor poder predictivo, quedarían los atributos tradicionales o clásicos, que evidencian un menor poder predictivo y son los usados en los M-Ts. En la

investigación para generar los modelos predictivos, fue utilizado el meta-algoritmo **Attribute Selected Classifier**, con un método de búsqueda y un evaluador, que selecciona los atributos más significativos antes de aplicar a los datos el algoritmo clasificador. ^{15,113}

Con pocos atributos, los más significativos, aquellos con mayor poder predictivo en el modelo, se obtiene una mejor clasificación y se evita el sobreajuste (*overfitting*), además, ayudan a reducir la complejidad de los modelos diseñados, mejorando su comprensión, así como el tiempo de procesamiento computacional. ^{114, 115,116}

El aprendizaje automático es parte fundamental en un proceso de análisis predictivo, ya que proporciona las técnicas de análisis de datos mediante las cuales se pueden descubrir relaciones entre variables que en un principio pueden parecer insignificantes, pero que tras la aplicación de estas técnicas puede descubrirse la trascendencia de las mismas, la minería de datos, que utiliza toda la información a diferencia de la estadística que trabaja con muestras de información y muestra que predictores, como la obesidad abdominal y la glucemia alterada en ayunas en personas no diabéticas, fueron más significativos que el IMC y la DM2. El investigador reitera que se descubrió información desconocida, implícita y potencialmente útil en los datos. ^{15,20,117}

Aprendizaje supervisado. Técnicas predictivas

Los algoritmos seleccionados, propios de Weka que implementan las funcionalidades necesarias para generar los modelos, se ejecutaron a los datos originales y se reejecutaron en los datos balanceados, como: el algoritmo de clasificación de reglas **JRip**, el meta algoritmo de clasificación **Attribute Selected Classifier**, usando como evaluador el **CfsSubsetEval**, el **BestFirst** como método de búsqueda y los algoritmos clasificadores de

árboles **J48** y el algoritmo de clasificación de funciones **Multilayer Perceptron**, una red neuronal artificial con retropropagación (*backpropagation*).¹⁵

En la extracción de reglas con el algoritmo **JRip**, una variante del algoritmo de cobertura secuencial de Cohen en 1995, que aprende de forma iterativa las reglas y elimina los puntos de datos cubiertos por la nueva regla, se sustituyeron los datos, en el archivo arff, por el valor de referencia usado en la atención médica en cada categoría codificada, la reglas se generaron en los datos de entrenamiento de la forma: SÍ-ENTONCES (**IF-THEN**), que consiste en una condición (también llamada antecedente) y una predicción, si se cumplen las condiciones, siempre que la condición se construya a partir de características inteligibles, entonces se hace la predicción.¹¹⁸ En esta investigación se predice si el paciente tiene riesgo de sufrir un evento cardiovascular, en caso contrario de no cumplirse las condiciones el paciente no se considera en riesgo, después del paréntesis se muestran los aciertos y errores en cada regla. (Anexo 6, tabla 5)

El algoritmo de clasificación de reglas **RIPPER** (*Repeated Incremental Pruning to Produce Error Reduction*), **JRip** en Weka genera reglas para la clase minoritaria, la clase de interés en la investigación, este realiza una poda incremental repetida para producir reducción de errores, seguido de una optimización global. Las reglas de decisión son probablemente los modelos de predicción más interpretables, dentro de los modelos interpretables. Lo nuevo en el aprendizaje automático es que las reglas de decisión se aprenden a través de un algoritmo.^{15,108,119}

Los meta-algoritmos de clasificación se basan en aplicar múltiples algoritmos para la identificación de clases, como algoritmos de selección de atributos y de clasificación

posterior, o de ejecución de múltiples algoritmos de clasificación para evaluar el mejor de todos ellos, mediante su ajuste paramétrico.^{15,11}

El meta-algoritmo de clasificación ***Attribute Selected Classifier*** combinó la selección de atributos junto con el algoritmo de clasificación **J 48**,¹⁵ un árbol de decisión, que también puede representarse como reglas.¹¹⁹ Se seleccionaron los atributos predictores más significativos, para crear nodos de forma descendente y recursiva; los nodos más próximos a la raíz son los mejores predictores, y a partir de ellos el proceso se repite para generar ramas en el árbol con nuevos nodos a partir de los datos restantes, el antecedente de hipertensión arterial (apphta) fue el rasgo más significativo, seguido por el cHDL y cLDL, la pcr-as y las cifras de tensión arterial sistólica, a partir de ellos se derivaron las reglas de predicción. Se muestra una fracción del árbol de decisión obtenido. (Anexo 6, figura 9)

Las reglas obtenidas con el algoritmo **J 48** (Anexo 6, tabla 6), que construye un árbol de decisión que busca descubrir inductivamente patrones predictivos a partir de los datos de entrenamiento, cada rama del árbol es una posible explicación al problema de predicción planteado (reglas de decisión) y cada una se interpreta leyendo la secuencia de nodos y sus condiciones a partir de la raíz y hasta llegar al nodo hoja.^{15,113}

Dentro de los modelos de predicción, los árboles de decisión (*Decision Trees*, en inglés) son los más utilizados por una serie de razones que los hacen especialmente atractivos como: la sencillez del modelo, la amplitud de implementaciones que existen, la rapidez de clasificación de nuevos patrones, la posibilidad de representarlos gráficamente aportando así una explicación de la división efectuada, la fácil interpretación, en caso de no ser excesivamente grandes y la posibilidad de obtener las reglas por las cuales se asignan cada una de las probabilidades.¹¹⁵

Las redes neuronales artificiales (*Artificial Neural Networks, ANN*), son un conjunto de algoritmos inspirados en el mecanismo de comunicación de la neurona biológica y han demostrado ser una buena aproximación a problemas donde el conocimiento es impreciso o variante en el tiempo. Su capacidad de aprender convierte a las redes neuronales en algoritmos adaptativos y elaborados a la vez, su adaptabilidad proviene de su capacidad de aprender de los datos, mientras que su elaboración se refiere a la complejidad de su estructura y funcionamiento ¹¹⁹

La arquitectura o topología de una red neuronal se define como la organización de las neuronas en distintas capas, así como los parámetros que afectan a la configuración de las neuronas, tales como las funciones de entrada o activación.¹²⁰

La arquitectura de la red neuronal, puede ser compleja, con múltiples redes profundas y una variedad de arquitecturas, donde pueden incluirse diferentes tipos de neuronas y conexiones; esta complejidad le permite capturar patrones y relaciones no lineales en los datos, que la hace muy efectiva para tareas como reconocimiento de imágenes, procesamiento del lenguaje natural, entre otras.^{119,120}

La estructura quedó conformada por las neuronas en la capa de entrada, que corresponden a los rasgos predictores, una capa oculta con once neuronas y dos neuronas en la capa de salida que corresponden a la clase decisión Sí y No (Anexo 6, figura 10), en las 500 etapas del proceso (*epochs*). El aprendizaje fue por la técnica de corrección de error, este fue disminuyendo de 0,0132428 hasta llegar a 0,0063552, con una tasa de aprendizaje (*learning rate*) de 0,3 y el momentum de 0,2.

Una regla para determinar la cantidad de capas ocultas es dividir el número de predictores en la capa de entrada entre el número de clases en la capa de salida, que el autor no utilizó porque se requería de mucha potencia de cálculo que no se disponía.

En el algoritmo **Multilayer Perceptron** el aprendizaje se produce por medio de la retropropagación, el algoritmo compara el resultado de la capa de salida con el resultado deseado y asume que el error en la unidad de salida se debe a errores en las unidades conectadas con ella y para corregirse realiza ajustes en los pesos asignados en la red desde la capa de salida hasta la capa de entrada, hacia atrás, de ahí el término de retropropagación.^{15,120} La función F utilizada fue la función sigmoide.

Una de las limitaciones para la implementación práctica de estos sistemas de inteligencia artificial, de tanto potencial como los modelos algorítmicos basados en redes neuronales, es que no hay suficiente comprensión de las reglas que generan, se muestran los nodos 0 y 1 de múltiples nodos, (Anexo 6, figura 11) dificultando la interpretación, comportándose como cajas negras (*black boxes*, en inglés); en la actualidad se ha avanzado en el campo de la Inteligencia Artificial Explicable (*Explainable Artificial Intelligence*, XAI) para solucionar ese problema. En la investigación las reglas de la red neuronal fueron extraídas con el *IBM SPSS*, coincidiendo con las reglas obtenidas en los modelos **JRip y J48**.^{12,120,121} Weka permite almacenar los modelos obtenidos y al suministrar una nueva base de datos, de otros pacientes con los mismos predictores, clasificar las instancias y, por tanto, utilizar estos modelos de tanto potencial predictivo.¹²

Etapas 5. Evaluación y validación de los modelos predictivos

Para evaluar la calidad de un modelo entrenado se utilizan distintas métricas (Anexo 4) cuantificando su rendimiento, lo que permite compararlos y elegir el mejor M-IA^{122,123}

La matriz de confusión, es la herramienta de visualización más empleada en el aprendizaje supervisado,¹¹⁵ ya que muestra el desempeño de un clasificador describiendo como se distribuyen los valores reales y las predicciones, no es una medida de rendimiento como tal, pero todas las métricas que aparecen a continuación se basan en los datos que contiene.¹²³

Los algoritmos clasificadores (Anexo 6, tabla 7) muestran que fueron efectivos para generar los modelos predictivos, al comparar las métricas en el test de entrenamiento, en la clase minoritaria Evento =SÍ, que es la clase de interés; el mejor desempeño se evidenció con el algoritmo **Multilayer Perceptron**, una red neuronal artificial con retropropagación.

El objetivo de un modelo de clasificación es predecir la clase de futuras instancias o en datos que no conoce, se necesita cerciorarse de que el modelo generaliza bien, es decir, que sea capaz de reconocer datos que difieran un poco de los que ha entrenado, para lo que se debe validar el modelo predictivo creado.^{89,114}

Para validar los modelos se realizó la validación cruzada de k particiones, que es uno de los métodos de validación más simples y comúnmente mencionados en la literatura, pues consiste en dividir los datos de manera aleatoria en k grupos del mismo tamaño. En la investigación k=10, k - 1 nueve de esos grupos serán utilizados como conjunto de entrenamiento y el grupo restante será empleado para realizar la evaluación.^{12,114,123}

Este proceso se repite k veces utilizando un grupo distinto como validación en cada iteración, genera k estimaciones cuyo promedio se emplea como estimación final. Esto sirve para confirmar que el resultado obtenido es independiente de los datos concretos con los que se ha realizado la prueba de clasificación y que, por tanto, el modelo generaliza bien y no se evidenció sobreajuste, que es lo contrario de la generalización.^{12,114}

Los resultados de la validación cruzada con 10 particiones (Anexo 6, tabla 8), se basaron en las métricas aplicadas para determinar el desempeño de los modelos, como el área bajo la curva ROC, que ha sido adoptada por la comunidad de investigadores en minería de datos, la Medida F que se considera como una media armónica que combina los valores de la precisión y de la sensibilidad o recall; la curva PR es la gráfica resultante de relacionar la precisión y el recall de un modelo y permite ver a partir de qué recall tenemos una degradación de la precisión, muy útil para datos desbalanceados.^{12,89,114,115}

El Coeficiente de Correlación de Matthews es una métrica no tan conocida, recomendada para datos desbalanceados, pero que tiene una gran importancia ya que engloba las cuatro entradas de la matriz de confusión y el resultado es alto, solo si su clasificador está funcionando bien tanto en los elementos negativos como en los positivos.¹¹⁵ WEKA cuenta con las siguientes opciones para el proceso de validación: utilizar el conjunto de entrenamiento, suministrar un conjunto de prueba, la validación cruzada e indicar un porcentaje de división donde se dividen los datos en un grupo de entrenamiento y otro de prueba, de acuerdo con un porcentaje indicado.^{12,113}

Una alternativa a los criterios de información, no disponible en Weka, suponiendo que se dispone de suficientes datos, es realizar una partición de la muestra disponible en tres partes: un conjunto de entrenamiento, otro de validación y un tercero de prueba, pero como la muestra disponible no era lo suficientemente grande, la anterior división podría conducir a submuestras excesivamente pequeñas.^{89,109}

En la investigación solo se disponía de una sola muestra, con una cantidad no muy elevada de datos para construir el modelo, por lo que no fue conveniente reservar parte de los mismos para la evaluación y se recomienda en la literatura usar la validación cruzada que

es una de las formas más consistentes de evaluar clasificadores, ya que los conjuntos se determinan de manera aleatoria y el error de la evaluación es muy bajo. Comparado con los demás métodos, es la validación más estricta para verificar la precisión de los modelos, esto asegura su capacidad de generalización. ^{12,109,113}

El mejor rendimiento, en la evaluación y la validación, se alcanzó con el algoritmo ***Multilayer Perceptron***, una red neuronal con capas ocultas y retropropagación. (Anexo 6, tabla 8)

El riesgo de sesgo de los modelos fue evaluado por el autor y de forma independiente por los tutores, como ha sido mencionado, con la herramienta de evaluación de sesgo de predicción PROBAT¹⁰⁰ utilizando la lista de verificación CHARMS.¹⁰¹ El formulario de extracción incluyó cuatro dominios: participantes, predictores, resultados y análisis estadístico (Anexo 7, figura 12), PROBAT es un método común de riesgo de sesgo para los M-Ts; aunque ha sido usado en modelos diseñados con IA^{3,100} evidencia limitaciones, por lo que en esta obra escrita se utilizó una declaración de TRIPOD-AI y se muestra su lista de verificación (checklist TRIPOD-AI) en la figura 13. (Anexo 6).

Estas herramienta permiten la evaluación de los modelos predictivos antes las dificultades encontradas en los mismos,³ los que evidenciaron un bajo riesgo de sesgo, aunque es importante señalar que ninguna herramienta por si sola puede garantizar una evaluación completa, la combinación de diferentes enfoques y métricas, junto con una revisión crítica del proceso de minería de datos es esencial para obtener una evaluación robusta y confiable, además, la interpretación debe hacerse considerando el contexto del problema y los objetivos de la investigación.

En la investigación fue utilizada la metodología KDD: un proceso estructurado y disciplinado que permite la resolución de un problema de investigación de manera eficaz y efectiva, junto a Weka, ambas entre las más populares, y más usadas; por tal motivo, a juicio del autor se evidenció el bajo riesgo de sesgo de los M-IAs.

Por último, el autor puede afirmar que, por los resultados alcanzados en la evaluación y validación, el bajo riesgo de sesgo, la baja preocupación referente a su aplicabilidad clínica y su replicación o reproducibilidad, los M-IAs son efectivos para respaldar su empleo en la predicción del riesgo de enfermedad cardiovascular en la atención primaria de salud.

3.1.3 Comparación del modelo predictivo implementado con las Tablas de Predicción del Riesgo Cardiovascular de la OMS de 2019

Implementación del modelo predictivo para su uso en la atención primaria de salud

El modelo predictivo para su aplicación en la atención primaria de salud fue implementado con las reglas generadas por el M-IA del algoritmo **JRip** y las reglas del M-IA del árbol de decisión **J 48**, se denominó **ModPred IA EnfCardiovasc_ aps 1.0**® en forma de una lista de decisión con 16 reglas predictoras y con los predictores más significativos identificados en colores (Anexo 6, figura 14)

La lista de decisión introduce un orden de reglas de decisión, que resuelven el problema de la superposición, las primeras son las de mayor poder predictivo, si se cumple la condición se usa la predicción de esa regla. En la investigación el paciente se considera en riesgo de un evento cardiovascular, de no cumplirse la condición pasamos a la siguiente regla y, así sucesivamente, si no se cumple ninguna, el paciente no se considera en riesgo de desarrollar un evento cardiovascular.¹²⁴

Los árboles de decisión son un modelo surgido en el ámbito del *Machine Learning* y de la Inteligencia Artificial que, partiendo de una base de datos, crea diagramas de construcciones lógicas que ayudan a resolver problemas.¹²² El algoritmo **J 48**, una implementación en Weka del algoritmo **C 4,5** de Ross Quinlan para mejorar el algoritmo **ID3** (*Induction Decision Tree*), es uno de los algoritmos más utilizados en problemas de clasificación, por su alto rendimiento y por su sencillez al momento de interpretar las reglas predictivas de sus modelos, construye el árbol de clasificación a partir del conjunto de entrenamiento y después de hacerlo utiliza una técnica de poda (*prunning*) basada en aplicar un test de hipótesis para deducir si merece la pena o no expandir dicha rama.

113,114,122

El refinamiento que introduce respecto al **ID3** es una medida alternativa llamada la ratio de ganancia, es decir, en cada nodo selecciona el atributo con mayor ratio de ganancia de información evitando así favorecer la elección de variables con un mayor número de valores.

113,114,122

Además, Weka tiene una interfaz gráfica de usuario (Anexo 6, figura 15) en 2D, que conduce en las tareas de minería de datos y son excelentes herramientas de visualización de datos que ayudan a entender los modelos predictivos diseñados con inteligencia artificial.¹⁵

La sensibilidad o recall (R) del M-IA implementado, que es la capacidad de detectar una condición, son los pacientes con riesgo de presentar una enfermedad cardiovascular, en esta investigación fue del 94,1 %, ($R = \text{verdaderos positivos} / \text{total de positivos} * 100$) y la especificidad (E), que es la capacidad de detectar los pacientes que no están en riesgo fue de 97 %. ($E = \text{verdaderos negativos} / \text{total de negativos} * 100$)

En Weka, el autor reitera su utilidad: los modelos generados pueden ser almacenados en archivos para, posteriormente, ser utilizados con nuevos datos y realizar predicciones.^{15,111}

Comparación del modelo predictivo implementado con las Tablas de predicción del Riesgo Cardiovascular de la OMS

La última etapa de la investigación fue la comparación del M-IA implementado con las tablas de la OMS, utilizando la base de datos original del estudio de cohorte prospectivo.

La tabla de predicción de la OMS de 2019 (Anexo 6, figura 16) comprende las edades de 40 a 79 años y el nivel de riesgo se considera bajo si es menor del 5 %, intermedio de 5 % pero menor de 10 % y alto de 10 % y superior.¹

La tabla 9 (Anexo 6) muestra algunas características de los modelos, la tabla de la OMS de 2019 es la única opción en regiones donde no están disponibles estudios de cohortes prospectivos; la ausencia de índices validación o discriminación tiene un efecto real en la sensibilidad, especificidad y exactitud predictiva de dichas tablas.¹²⁵

Basado en el nivel de evidencia de estudios pronósticos, la tabla de la OMS tiene un nivel de evidencia II y el M-IA implementado un nivel de evidencia I, ya que los puntajes de riesgo, basados en cohortes poblacionales, son herramientas muy útiles en la práctica clínica y permiten clasificar a las personas según grupos de riesgo y, por tanto, priorizar las intervenciones de prevención en sujetos con un riesgo cardiovascular más elevado.^{8,63,126}

La tabla de la OMS frecuentemente clasifica a la mayoría de la población en la categoría de riesgo bajo, la prevalencia de esta categoría fue del 89,3 % en Jamaica y del 89,7 % en Cuba. Es necesario destacar que estas predicciones se hicieron con la tabla de la OMS de 2007 y en las tablas predictivas de 2019 se redujo el nivel de riesgo en todas las categorías,

pues en la búsqueda bibliográfica no se encontraron predicciones realizadas en poblaciones cubanas con las tablas de 2019.^{12, 125,127}

El modelo implementado incluye más variables predictivas, las modificables donde se puede intervenir para su control desde la atención primaria de salud. Investigaciones realizadas reportan que la alta prevalencia de los factores de riesgo cardiovascular no se corresponde con la categorización del riesgo en la predicción con las tablas OMS y demás M-Ts¹²⁸

En la predicción del riesgo de enfermedad cardiovascular entre el M-IA implementado y la tabla de la OMS 2019, fue utilizada la base de datos original del estudio de cohorte prospectivo, sin balancear, y señalando que hubo una diferencia de ± 5 años, entre los modelos, ya que la edad del modelo predictivo abarcó de 35 a 74 años y la tabla de la OMS de 40 a 79 años, como se dijo anteriormente no se encontraron predicciones del riesgo cardiovascular con las tablas de la OMS 2019 en poblaciones cubanas.

La tabla de la OMS clasificó casi el 26 % de la población como riesgo intermedio, (Anexo 6, tabla 10) una categoría de riesgo, como su nombre lo indica, es un punto medio entre el riesgo alto y bajo y se recomienda aplicar otros métodos para reclasificarlo, pues la predicción del riesgo cardiovascular en este grupo no es fiable, pero en la mayoría de las ocasiones los métodos no están disponibles en la atención primaria de salud, y por consiguiente, no permiten reclasificar el riesgo, y por tanto, no se realizan acciones preventivas o terapéuticas eficazmente, y generalmente queda en “inercia” y aporta significativamente, junto con el riesgo bajo, a la morbilidad y mortalidad cardiovascular.^{11,64}

La predicción de la tabla de la OMS 2019 (Anexo 6, tabla 10) no se diferencia de las predicciones obtenidas anteriormente en la población cubana. Fue clasificado el 76,4 % de

la población en las categorías de riesgo bajo e intermedio, que al ser mayoritario es el grupo que más aporta a la morbilidad y mortalidad cardiovascular, como se mencionó .¹²

Se comparó la predicción obtenida del M-IA como riesgo = Sí y la predicción de las tablas de la OMS en esos pacientes, utilizando la base de datos original del estudio de cohorte prospectivo. (Anexo 6, tabla 11) El M-IA predice en 519 personas como riesgo = Sí de desarrollar una enfermedad cardiovascular en los próximos 10 años, de estas personas, la tabla de la OMS predijo como riesgo intermedio 173 personas (33,3 %) y como bajo 184 (35,5 %), en total 357 pacientes (68,8 %) en las categorías intermedio y bajo.

Los lípidos y sus lipoproteínas han sido mediadores y marcadores de enfermedad cardiovascular, pues una concentración elevada de triglicéridos (TG) y bajas de cHDL producen un incremento de las LDL pequeñas y densas, que son las más aterogénicas y son potentes marcadores de dislipemia aterogénica, aterosclerosis subclínica y por consiguiente de enfermedad cardiovascular.¹²⁹

Para valorar las proporciones y la aterogenicidad de los lípidos en la sangre se han ideado unas fórmulas llamadas índices aterogénicos, que son un conjunto de indicadores que relacionan el colesterol total, el cHDL y el cLDL entre sí o con los triglicéridos y dan una estimación aproximada de lo perjudicial o no que es el perfil lipídico del paciente, por tanto, son marcadores pronósticos tempranos de ECV. ^{129,130, 131}

El índice aterogénico del plasma (IAP) resultante del cociente TG/cHDL, es de gran utilidad ya que su uso en la atención primaria es un indicador indirecto del tamaño de las partículas LDL 5,11 a 14, oxidadas, pequeñas y densas que son las más aterogénicas, a mayor índice aterogénico, mayores son las posibilidades de que se forme una placa de ateroma. Este

índice aterogénico es un potente predictor de aterosclerosis, incluida la subclínica y, por tanto, de enfermedad cardiovascular.^{129,130,131,132}

En los pacientes que las tablas de la OMS 2019 predicen como riesgo bajo o intermedio y el M-IA predijo como riesgo = Sí fue calculado el índice aterogénico TG/cHDL, y además fue revisada la concentración de pcr-as entre 1-3 mg/dL y mayor de 3 mg/dL en esas mismas categorías de riesgo de las tablas de la OMS.

En 109 pacientes (54,3 %) clasificados como riesgo bajo y en 65 pacientes (37,6 %) con riesgo intermedio, en las tablas de la OMS, presentaron un índice TG/cHDL mayor de 2, sospechando eran portadores de dislipemia aterogénica, y por tanto de AS, por ello el IAP TG/cHDL, en una cifra única, ha sido propuesto como marcador del potencial aterogénico del plasma, que permite predecir el daño aterosclerótico de riesgo de síndrome metabólico, dislipemia aterogénica, enfermedades cardiovasculares y otras entidades clínicas.¹³³

En la mayoría de los pacientes con probable DA, según el índice aterogénico TG/cHDL, fueron observadas concentraciones de pcr-as mayores de 1 mg/dL y en los restantes sin sospecha de dislipemia aterogénica mostraron concentraciones de pcr-as entre 1-3 mg/dL y mayor de 3 mg/dL, 131 pacientes, (36,7%), evidenciando la estrecha relación entre la aterosclerosis que es el proceso subyacente de la enfermedad cardiovascular o aterotrombótica y la inflamación crónica o de bajo grado, que se identifica como concentraciones de pcr-as entre 3 y 10 mg/dL, y las concentraciones entre 1 y 3 mg/dL que son consideradas por la *American Heart Association* como un riesgo moderado de desarrollar una enfermedad cardiovascular.^{104,105}

En concordancia con estos resultados se considera que de los 357 pacientes que la tabla OMS predijo como riesgo bajo o intermedio en 305 (85,4 %), los ya señalados por

considerarlos con sospecha de dislipemia aterogénica y concentraciones de pcr-as mayores de 1 mg/dL, este modelo infra estima el riesgo cardiovascular. Los M-Ts disponibles, incluidas las tablas de la OMS no utilizan en sus predicciones, ningún marcador de inflamación a pesar que se considera actualmente la aterosclerosis como una enfermedad inmuno inflamatoria.¹³⁴

La inflamación es reconocida en la aterosclerosis, asumiendo la percepción de que la enfermedad cardiovascular puede considerarse actualmente como una enfermedad inmuno inflamatoria y sus complicaciones son desencadenadas por mecanismos complejos de dislipemia e inflamación. Ambos mecanismos están influenciados por los factores de riesgo cardiovascular, en efecto, la proteína c reactiva puede contribuir con información para una predicción más fiable del riesgo cardiovascular.^{104,105, 134} El autor sugiere valorar las concentraciones de pcr-as mayores de 1 mg/dL en pacientes clasificados por las reglas de decisión como Riesgo = No, dado el rol que desempeña esta proteína en todas las fases del proceso aterosclerótico y tratar intensivamente otros factores de riesgo modificables presentes en esos pacientes.

Del mismo modo, fueron comparados los pacientes que el M-IA predice como riesgo = NO y la correspondiente predicción que realiza en esos pacientes las tablas de la OMS, (Anexo 6, tabla11); el M-IA predice 1114 pacientes sin riesgo de enfermedad cardiovascular, de los 225 pacientes clasificados como riesgo alto y en los 245 clasificados como intermedio, por las tablas de la OMS, en 152 (32,3 %), solo se evidenciaba como factor de riesgo una edad de 64 y más años.

La tabla predictiva de la OMS (Anexo 6, figura16) nos muestra que en personas no diabéticas con edad mayor de 64 años solo clasifica el riesgo como intermedio o alto y en

pacientes diabéticos esta clasificación es a partir de los 47 años. La edad en personas mayores es un tema controvertido en la literatura, pues hay opiniones que el riesgo no se debe tratar cuando lo causa solo la edad y otros señalan que muchos tratamientos preventivos son eficaces en edades avanzadas, en cuanto a posponer la morbilidad y la mortalidad.¹³⁵

El autor considera la segunda opinión, el riesgo según la edad se debe tratar de forma holística, incluyendo los predictores tradicionales y los noveles involucrados como predictores o biomarcadores de riesgo en el proceso aterosclerótico; el envejecimiento y la inflamación son piezas claves en la patogénesis de la aterosclerosis y, por tanto, de las enfermedades cardiovasculares. El envejecimiento está caracterizado por una inflamación crónica, desregulada y sistémica, además, por una respuesta inmune disfuncional que incrementa el daño cardiovascular, estas noxas promueven la disfunción endotelial, la infiltración inmune, la deposición de células espumosas y la calcificación, lo que resulta en la formación de la placa aterosclerótica. Avances recientes sugieren que la terapia antiinflamatoria, teniendo como blanco la pcr-as, reduce el riesgo cardiovascular, para lograr su control, evitando la rotura de la placa vulnerable y así sus complicaciones.¹³⁶

En resumen, el modelo implementado con datos de una cohorte prospectiva de una fracción de la población cubana, basado en técnicas de inteligencia artificial, predice mejor que las tablas de la OMS 2019, que sobreestiman el riesgo en 152 pacientes por edad o infraestiman el riesgo en 305 individuos portadores de DA e inflamación crónica como ha sido mostrado. Las tablas de la OMS 2019, en la región del Caribe, están basadas en un puntaje de riesgo sobre el análisis de la carga de morbimortalidad en la región,¹ nueve países muestran tasas medias de mortalidad superiores a las de Cuba, de 181,5 hasta

428,7 por 100 000 habitantes y son países con sistemas de salud diferentes al de Cuba. (Anexo 1, gráfico 1)²⁶

A diferencia de las tablas de la OMS 2019, el M-IA implementado es capaz de determinar la inflamación de bajo grado que se observa en la DA, el SM y sospechar la aterosclerosis subclínica, además, identifica los marcadores de carga alostática y permite sospechar la RI.

1,46,52,53

Desde otra perspectiva, la dialéctica materialista nos ofrece leyes generales, que sustentan el fundamento de toda explicación de la realidad, son objetivas, independientes de la naturaleza humana. Una de esas leyes es la Ley de transición de la cantidad a la cualidad y viceversa, enunciadas por Marx y Engels. Engels en el Anti-Dühring, menciona que “todos los objetos de la naturaleza poseen características mensurables, por lo que su esencia, su cualidad, es inseparable de los aspectos cuantitativos, cuando una cosa pasa de poseer una cualidad a poseer otra hablamos de salto cualitativo”.¹³⁷

La cantidad y la calidad no pueden separarse una de la otra ni ambas del objeto, solo por abstracción mental; sin embargo, un cambio constante de la cantidad, que traspase ciertos límites ocasiona un cambio de la calidad de tal manera que el objeto deja de ser lo que fue para ser otra cosa, así los cambios cuantitativos se tornan en cambios cualitativos.^{137,138}

Como todo movimiento es el resultado de la lucha de elementos contrarios, el salto cualitativo supone la resolución de una contradicción, que da lugar a una nueva realidad, que representa un avance en el desarrollo. El salto cualitativo no supone el mero cambio de una cualidad por otra, sino por otra que supera, de alguna manera, a la anterior.^{137,138}

En conclusión, basados en su diseño, la comparación de los modelos y en la ley dialéctica, que pone al descubierto los rasgos esenciales del desarrollo y permite preveer

científicamente los acontecimientos y muestra un avance al desarrollo, un salto cualitativo entre los modelos, el autor considera que el M-IA obtenido es más eficaz en la predicción del riesgo cardiovascular que las tablas de la OMS, reafirmando así, que los modelos diseñados con técnicas de inteligencia artificial muestran mejor rendimiento que los modelos tradicionales.^{3,18,19,20}

3.2 Rol de los nuevos predictores en la aterosclerosis

En este acápite se explican las relaciones esenciales y cualidades fundamentales de los nuevos predictores incorporados al M-IA, que posibilitan la interpretación conceptual de los datos empíricos e ir más allá de las características fenoménicas y superficiales de la realidad.

Los predictores de riesgo incluidos en el M-IA diseñado e implementado su uso en la atención primaria de salud son: la proteína c reactiva de alta sensibilidad o ultrasensible, la obesidad abdominal determinada con la medición de la circunferencia abdominal, la glucemia alterada en ayunas en no diabéticos y el nivel de triglicéridos, este último usado en los M-Ts: el nuevo algoritmo de Framingham D'Agostino de 2008⁷¹ y el puntaje PROCAM (*Prospective Cardiovascular Munster Study*),⁷⁴ que están entre los 14 M-Ts más usados a nivel mundial.¹²

El predictor que el autor considera más relevante en el M-IA diseñado es la proteína c reactiva de alta sensibilidad. La evidencia epidemiológica generada en los últimos 20 años ha demostrado que la proteína c reactiva es un actor activo e importante en la enfermedad cardiovascular aterosclerótica. Cuando se habla de enfermedad cardiovascular, la aterosclerosis y la inflamación tienen una relación causa efecto, los eventos

cardiovasculares adversos causados por la ruptura de la placa aterosclerótica ocurren debido a la elevación mantenida de los marcadores inflamatorios.¹⁰⁵

La pcr-as es un biomarcador útil en la predicción del riesgo cardiovascular, capaz de identificar la inflamación crónica o de bajo grado, también denominada meta inflamación o lipidoinflamación; además, la evidencia reciente lo señala como un predictor de riesgo cardiovascular y un objetivo o blanco terapéutico en el control de este, la prevención y tratamiento de la enfermedad cardiovascular aterosclerótica.^{104,105,139,140}

La inflamación es un factor que participa en el desarrollo de la aterosclerosis y la aparición de eventos cardiovasculares adversos. Además, la inflamación persistente induce vulnerabilidad en la placa aterosclerótica, haciéndola propensa a la erosión y la ruptura.⁴ El biomarcador de inflamación más estudiado en la enfermedad cardiovascular es la pcr-as.¹⁰⁵ Están involucradas otras citocinas, como el factor de necrosis tumoral alfa (TNF- α), interferón γ (IFN- γ), también llamado interferón inmunitario o de tipo II, la interleuquina 1 beta (IL-1 β), la interleucina-10 (IL-10), la interleucina-6 (IL-6), y el factor de crecimiento transformante β (TGF- β).¹⁴⁰

Ya en 1998 la investigación de Ridker y colaboradores¹⁴¹ documentó que el valor inicial de la proteína c reactiva predijo el riesgo de enfermedades cardiovasculares en personas con cifras bajas de colesterol total y con un índice CT/cHDL bajo, es decir con bajo riesgo cardiovascular, además, la determinación de pcr-as y cLDL juntos proporcionó mejor información pronóstica que la de cualquiera de ellos por separado. Se retomó el valor y la actualidad de los hallazgos de esta investigación en el tiempo.^{105,141}

La evidencia proveniente de estudios epidemiológicos ha demostrado una fuerte asociación entre las concentraciones elevadas de pcr en suero o plasma y la incidencia de eventos cardiovasculares. **(Nivel de evidencia A)** ^{104,105,140,142,143,144,145,146}

La inflamación crónica o de bajo grado, detectada por pcr-as, o en su defecto, aunque no tan sensible con pcr cualitativa, que detecta un valor mínimo de 6 mg/dL¹⁴³ se correlaciona con la obesidad, incluida la obesidad abdominal, la resistencia a la insulina, disfunción endotelial, la glucemia alterada en ayunas, la DM2, el síndrome metabólico, la dislipemia metabólica o aterogénica y la presencia de placas de ateromas en la aterosclerosis subclínica.^{105,133,134,140,146} El autor considera el gran valor aditivo de la pcr-as al incluirlo en los M-IAs, que evidentemente mejora la predicción del riesgo cardiovascular.

(Recomendación clase I, nivel de evidencia A)

No se encontraron investigaciones sobre prevalencia de la pcr en poblaciones abiertas cubanas, solo un estudio donde no se aprecian diferencias según el género;¹⁴⁷ en la cohorte el 21,5 % tenía valores superiores de 3 mg/dL, en estudio de una población mexicana, el 31 % evidenciaba inflamación crónica o de bajo grado¹⁰⁴; se han estimado valores normales en poblaciones de referencia donde entre el 70 y 90 % de la población tenían valores < 3 mg / dl, donde una elevación entre 3 y 10 mg/dL se considera inflamación de bajo grado y entre 1 y 3 mg/ dl, un riesgo cardiovascular moderado, lo que en términos de aplicabilidad clínica se ha relacionado con un mayor riesgo cardiovascular y con el desarrollo de condiciones clínicas futuras.¹⁰⁵ La aplicación del M-IA puede contribuir a determinar la prevalencia de la proteína c reactiva en poblaciones cubanas.

En la obesidad abdominal la cantidad de tejido adiposo visceral se correlaciona con un perfil de riesgo metabólico gravemente alterado, por lo que, aunque la obesidad general,

determinada con el IMC, incrementa el riesgo de enfermedades crónicas, parece claro que los pacientes con obesidad visceral forman el subgrupo de individuos con las alteraciones más graves del metabolismo.¹³⁰

Se ha observado la existencia de un fenotipo correspondiente a individuos con peso normal, pero metabólicamente obesos (en inglés, *metabolically obese normal-weight* [MONW]); es decir, tienen un IMC normal, no obstante presentar las alteraciones metabólicas de los pacientes obesos, como resistencia a la insulina, adiposidad central, bajas cifras de cHDL, altas concentraciones de triglicéridos e hipertensión arterial. Al mismo tiempo, existen los que se han denominado obesos metabólicamente sanos (*metabolically healthy obese* [MHO]), que tienen un IMC ≥ 30 sin las alteraciones metabólicas típicas de los individuos obesos anteriormente mencionadas.¹⁴⁶

En los últimos años se ha otorgado a la obesidad abdominal un papel trascendental en la fisiopatología del riesgo cardiovascular y metabólico, relacionadas con la generación de un estado pro inflamatorio y su relación bidireccional con la RI. Durante el proceso inflamatorio se modifican las funciones de casi todos los sistemas de órganos y en ciertos tipos de inflamación, la respuesta adaptativa es quien origina y perpetúa el proceso inflamatorio, existe una correlación entre la proteína c reactiva, la inflamación de bajo grado y la resistencia a la acción de la insulina en las personas obesas. El tejido adiposo desempeña un papel fundamental sobre las cascadas inflamatorias, procoagulantes, antifibrinolíticas y vasoactivas, lo que sugiere una influencia directa sobre el proceso inflamatorio sistémico crónico o de bajo grado.^{139,140,147}

Los trastornos metabólicos que acompañan a la obesidad abdominal se vinculan con la resistencia a la insulina, donde se produce una hiperinsulinemia compensatoria por una

deficiente acción de la insulina sobre sus células diana, que al principio mantiene las concentraciones normales de glucosa en sangre, pero que a largo plazo conlleva a una disfunción de las células β del páncreas, la cual incrementa la llegada de ácidos grasos no esterificados al hígado desde la grasa abdominal y aumenta la síntesis de triglicéridos en ese órgano. ^{147,148,149}

Como consecuencia, se produce esteatosis hepática y se eleva la liberación sanguínea de las VLDL. Estas VLDL, ricas en triglicéridos, producen hipertrigliceridemia. La hipertrigliceridemia se nutre de la baja actividad de la lipoproteína lipasa, enzima endotelial que degrada a las VLDL y los quilomicrones. ¹⁵⁰

Este aumento de triglicéridos en sangre repercute en otras lipoproteínas y aumenta las cLDL pequeñas y densas, capaces de infiltrar más fácilmente la íntima arterial y convertirse en células espumosas, además de producir una disminución de las cHDL. Esta tríada dislipidémica aterogénica, caracterizada por hipertrigliceridemia, LDL pequeñas densas y cHDL disminuidas, asociada a un estado proinflamatorio y protrombótico, aumenta el riesgo cardiovascular de los pacientes afectados. ^{148,149,150}

Entre los 20 principales factores de riesgo mundiales de años de vida perdidos en 2040, resultaron tres riesgos metabólicos: la presión arterial alta, el IMC alto y la glucosa plasmática en ayunas alta. Sobre la base de estos y otros factores de riesgo establecidos o factores emergentes, como el cHDL bajo, los triglicéridos y la pcr-as altos, por tanto, el concepto de salud metabólica está atrayendo mucha atención en la comunidad científica.

60

En la investigación la obesidad abdominal y la glucemia alterada en ayunas fueron predictores más significativos que el IMC y la DM2, reafirmando como la minería de datos

es capaz de descubrir información potencialmente útil e implícita en ellos y, por consiguiente, una predicción más fiable del paciente en riesgo de ECV¹⁵

Debido a que, los depósitos de grasa visceral son más resistentes a la insulina que la grasa subcutánea, los metabolitos de los ácidos grasos dentro de las células contribuyen a la resistencia a la insulina mediante mecanismos complejos que implican trastornos en la transducción de las señales de la insulina,¹⁴⁹ incluso si el IMC se encuentra dentro del rango normal, ya que este no es capaz de proporcionar información sobre la grasa visceral, sobre todo cuando se ubica alrededor de órganos vitales, por tanto, se puede seguir empleando el IMC, por ser un método rápido, de fácil de aplicación y bajo coste, pero por las limitaciones ya mencionadas, se recomiendan usar también índices alternativos para identificar la adiposidad de tipo central como predictor aceptado de riesgo cardiometabólico: el índice alométrico de forma corporal, el índice de cadera/altura, la determinación de la circunferencia abdominal, entre otros.^{140,151}

La glucemia alterada en ayunas, aunque se ha englobado en la definición de “prediabetes”, término controversial, ha sido asociada a enfermedades actuales y futuras como: la diabetes mellitus tipo 2, enfermedades cardiovasculares, enfermedad de hígado graso no alcohólico, enfermedad renal crónica, cáncer, neuropatías y demencias.¹⁵²

La resistencia a la insulina se produce por una interacción de una predisposición genética con factores ambientales o hábitos del estilo de vida que permiten que la resistencia a la insulina o la predisposición genética prosperen. La RI está relacionada íntimamente con la obesidad abdominal, el nivel de insulina aumenta mucho antes del diagnóstico de DM2, se produce una hiperinsulinemia compensadora que en sus inicios mantiene la glucemia en el rango normal o alterada en ayunas.^{148,149,150,152}

En relación a los triglicéridos y la proteína c reactiva de alta sensibilidad, en el estudio PESA, donde fue determinada la presencia de aterosclerosis subclínica por técnicas de imagen, se determinó en personas con riesgo cardiovascular bajo o intermedio que la hipertrigliceridemia estaba asociada a aterosclerosis subclínica e inflamación vascular de bajo grado, aun cuando tenían niveles de cLDL considerados normales.⁴¹

Estudios prospectivos demuestran que pequeños aumentos de la concentración de la proteína c reactiva circulante son predictivos para episodios cardiovasculares, hay pocos factores que modifiquen sus niveles de producción. Se consideran normales valores inferiores a 1 mg/dL y su determinación con la técnica turbidimétrica puede considerarse un biomarcador de inflamación crónica o de bajo grado.^{137,138}

La proteína c reactiva tiene un papel importante en la fisiopatología de la enfermedad aterosclerótica pues participa activamente en la formación, progresión y ruptura de la placa de ateroma y es un marcador bioquímico de inflamación subclínica de gran relevancia.¹³⁹

Está demostrado que la inflamación es un proceso crucial en el desarrollo del ateroma y, por ende, de la ECV que se asocia con DM 2, resistencia a la insulina, obesidad central y dislipidemia, se acepta, entonces, que el SM es un proceso inflamatorio asociado con valores plasmáticos elevados de proteína c reactiva, interleucina 6 (IL-6) y factor de necrosis tumoral alfa (TNF- α), que incrementan la producción hepática de proteína c reactiva. Un estudio realizado en Argentina, evidenció niveles elevados de proteína c reactiva con cifras de tensión arterial \geq 130/85 mm Hg, glucemia \geq 100 mg/dL y la concentración de pcr-as se incrementó a medida que aumentaba el número de componentes del síndrome metabólico.¹³⁹

Otros estudios encontraron una correlación positiva entre las concentraciones séricas elevadas de pcr-as y los niveles de resistencia a la insulina, independiente si se presentaba o no obesidad abdominal. ¹³⁹

Es necesario resaltar que, la inflamación puede activar tanto el sistema inmune innato como el adaptativo, e iniciar una cascada de citocinas proinflamatorias, como IL-1, IL-6 y TNF- α ,¹⁴⁰ mediante el transcriptor nuclear kappa beta, también la proteína c reactiva activa la vía alterna del complemento con la producción de componentes trombogénicos, dañando los vasos sanguíneos y las membranas celulares a nivel de diversos tejidos, sobre todo el adipocito y la célula hepática, que a largo plazo dañan los receptores de la insulina, haciéndolos menos sensibles a esta. ¹⁴¹

En la revisión bibliográfica no se hallaron estudios, con diseños apropiados, que nos permitieran comparaciones, sobre los niveles de pcr-as en poblaciones cubanas.

Por último, la resistencia a la insulina juega un papel fundamental en el desarrollo de alteraciones del metabolismo lipídico, que conlleva a un aumento en la concentración de TG, tanto en estados de ayunas como postprandiales, lo que contribuye a una menor formación de cHDL y a la formación de partículas de LDL pequeñas densas. Todo este conjunto de alteraciones termina en el desarrollo de dislipidemia aterogénica, recientemente denominada en la bibliografía dislipemia metabólica, típica en estados de resistencia a la insulina, obesidad, síndrome metabólico, diabetes mellitus tipo 2, aterosclerosis subclínica y, por consiguiente, enfermedades cardiovasculares. ¹⁴²

De modo que, la resistencia a la insulina ha sido asociada con la hipertensión arterial. Cerca de 50 % de los pacientes hipertensos presentaron hiperinsulinemia e intolerancia a la glucosa, mientras que al menos el 80 % de los diabéticos tipo 2 son hipertensos asociado

a la resistencia a la insulina. Del mismo modo, la integridad del epitelio es considerada un elemento fundamental en la salud vascular, siendo el óxido nítrico un potente vasodilatador endógeno, su reducción es un signo de disfunción endotelial, que contribuye a la hipertensión, también a la aterosclerosis y es causada por la disfunción endotelial.¹⁴³

En resumen, la inflamación crónica o de bajo grado, determinada con la pcr-as, considera el autor podría ser el “puente” que relaciona la resistencia a la insulina persistente, la disfunción endotelial, la dislipemia aterogénica o metabólica, el síndrome metabólico, la aterosclerosis subclínica, en la génesis de la iniciación y progresión de la aterosclerosis, en todos sus estadios,^{144,145} El autor opina que además de un biomarcador de riesgo, podría considerarse un predictor de riesgo, un objetivo o blanco terapéutico, útil para monitorizar o controlar la respuesta al tratamiento, aunque serían necesarias nuevas investigaciones para confirmar si es el *trigger* o gatillo fundamental que desencadena y perpetua el proceso aterosclerótico.

A juicio del autor los predictores añadidos al modelo obtenido, relacionados con la fisiopatología de la aterosclerosis según evidencias recientes y no considerados en los M-Ts actuales, podrían mejorar la predicción del riesgo de enfermedad cardiovascular, ya que su detección e intervención oportuna podrían contribuir al control del riesgo y a reducir la morbilidad y mortalidad por enfermedades cardiovasculares.

Es preciso insistir en el rol de la proteína c reactiva, en la enfermedad cardiovascular aterosclerótica, de preferencia la de alta sensibilidad, en la predicción del riesgo cardiovascular y, por consiguiente, que se incluya en los M-IAs recientes de predicción del riesgo cardiovascular y la evidencia confirma que mejora significativamente la predicción

del riesgo en comparación con los M-Ts. **(Recomendación clase I, nivel de evidencia A)**^{105,142,143,144,145}

Prueba de ello, la puntuación de riesgo de Reynold que incluyó la proteína c reactiva, mejoró significativamente la predicción del riesgo cardiovascular global al compararlos con modelos tradicionales,^{105,141} en otras investigaciones en pacientes con riesgo cardiovascular alto se observó correlación positiva con los niveles de pcr-as; otro estudio concluyó que niveles de pcr-as ≥ 1 mg/dL en pacientes obesos se correlacionó con la obesidad, incluida la abdominal, aumento de grosor de la íntima media carotídeas y placas de ateroma; otra investigación en niños y adolescentes apoyan la idea de que la asociación entre la obesidad y la pcr elevada se puede producir a cualquier edad, ya que no depende de esta, sino de los mecanismos de la inflamación crónica de bajo grado que son desencadenados como consecuencia de padecer obesidad.¹³⁹

En Cuba, según la revisión de la literatura no se encontraron M-Ts aplicados que incluyan como predictor la pcr-as, a pesar del valor significativo de la inflamación en la enfermedad cardiovascular aterosclerótica, ni modelos diseñados con inteligencia artificial.

Al mismo tiempo, se han desarrollados M-IAs, el algoritmo ML EN-PESA (*Progression of Early Subclinical Atherosclerosis*), basado en una investigación prospectiva que incluyó hombres y mujeres asintomáticos con edades entre 40 a 54 años, con 115 variables como potenciales predictores de aterosclerosis subclínica. Dentro de los marcadores en sangre, los más significativos fueron: colesterol total, cHDL y cLDL, así como, la hemoglobina A1c y en menor rango la glucosa, los triglicéridos y la proteína c reactiva de alta sensibilidad. El autor opina que la pcr-as no alcanzó un rango superior en la predicción porque la

investigación incluyó una población joven, asintomática, con características étnicas similares, un nivel socioeconómico y estilos de vida homogéneos.¹⁵¹

Caí Yue y colaboradores realizaron una revisión sistemática que incluyó 486 modelos diseñados con inteligencia artificial que predecían el riesgo de enfermedad cardiovascular en poblaciones generales y especiales, con el objetivo de evaluar el riesgo de sesgo de predicción con la herramienta PROBAST^{100,101} y desarrollar una puntuación de validación independiente (*IVS*, por sus siglas en inglés), que comprendía cinco pasos con cinco elementos de puntuación de la siguiente manera: transparencia de los modelos, rendimiento, viabilidad de la reproducción, riesgo de reproducción e implicación clínica de manera secuencial para la evaluación de la replicabilidad de los M-IAs, que pudiera contribuir a la validación externa independiente.³

La puntuación de validación independiente diseñada refleja las limitaciones que pueden presentar los M-IAs.³ Estos investigadores dividieron los modelos como de desarrollo (*development*) y validación (*independent validation*), este último cuando se ha realizado una validación externa independiente. Merece la pena resaltar que solo el 4,1 % de los M-IAs evaluados incluían las proteínas como predictores, todos los modelos tenían un alto riesgo de sesgo según PROBAST y la puntuación de validación independiente. Los autores recomiendan 10 modelos predictivos diseñados con inteligencia artificial.³

De manera que, siete de esos M-IAs usados en la predicción del riesgo de enfermedad cardiovascular, en la investigación de Jiang y colaboradores, evidenciaron una moderada a excelente discriminación, una cohorte con una media de seguimiento de $\pm 5,17$ años, una muestra de 1508 individuos de la etnia Kazakh, en China, sin historia de enfermedad cardiovascular a la inclusión, de los cuales 203 (13,5 %) desarrollaron un evento

cardiovascular, 22 variables fueron incluidas en el modelo de desarrollo, como: características sociodemográficas, historia médica, citocinas e índices aterogénicos, las citocinas inflamatorias como la proteína c reactiva de alta sensibilidad y la IL-6, fueron identificados como potentes predictores de enfermedad cardiovascular.¹⁵²

El autor resalta, la investigación de Jiang y colaboradores¹⁵² como muy relevante al ser una de las recomendadas en la revisión sistemática de los M-IAs realizada por Caí Y³. Su diseño con técnicas de inteligencia artificial y el principal resultado coinciden con esta investigación.¹¹³

Paralelamente, se han desarrollados M-IAs que incluyen la pcr-as, en modelo basado en árbol de decisión, con datos de una cohorte donde se evidenció que la glucemia alterada, la presión arterial sistólica, la hipertensión arterial y el colesterol total fueron predictores del nivel de pcr-as¹⁵³ y cada vez es más frecuente su uso en modelos ML para la predicción del riesgo cardiovascular.^{151,152,153,154} El M-IA implementado incluye predictores no incluidos en los M-Ts, con un predictor clave como la pcr-as. Debe recordarse que actualmente la enfermedad cardiovascular es considerada una enfermedad inflamatoria, y por tanto, la inflamación de bajo grado es un elemento clave y es preciso insistir que la proteína c reactiva es un potente predictor para identificar a las personas en riesgo.^{142,143,144,145,155} Niveles iniciales de pcr pueden predecir el riesgo en un periodo de 30 años, junto al cLDL y la lipoproteína (a) y su inclusión en los modelos mejora la predicción del riesgo cardiovascular.¹⁵⁶

3.3 Implicaciones del modelo predictivo en la atención primaria de salud

A consideración del autor el M-IA implementado supera las limitaciones y, por tanto, se diferencia de los M-Ts utilizados en la predicción del riesgo cardiovascular en poblaciones

cubanas y aporta una precisión más fiable del riesgo cardiovascular que los M-Ts disponibles. Primero se describirán las implicaciones de su implementación en la atención primaria en Cuba, ya que se modelo de atención de salud está orientado fundamentalmente a la promoción y prevención de salud ¹⁵⁷ y posteriormente se hará referencia a los aportes que hace el M-IA propuesto a la predicción del riesgo cardiovascular a nivel global.

El M-IA implementado basado en reglas de decisión es comprensible, los modelos basados en reglas de decisión son los más comprensibles dentro de los modelos comprensibles, por ende, su fácil uso en la práctica médica, por otro lado, la diseminación e implementación efectiva del M-IA, va precedida de una capacitación donde se actualiza sobre el riesgo cardiovascular, los nuevos predictores, las ventajas y las diferencias con los M-Ts disponibles, así como, las recomendaciones del autor para su aplicación en la práctica médica. (Anexo 6, figura 15)

De modo que, las reglas de decisión no se superponen, no son perfectas, pero realizan un excelente trabajo, al ampliar su implementación, en condiciones del mundo real en la atención primaria, las reglas pueden mejorarse, que no fue objetivo de esta investigación, las reglas pueden interpretarse en cada paciente según el juicio clínico del médico. El autor sugiere algunas recomendaciones para su empleo en la práctica médica, hay que resaltar, que además de la predicción, son una guía de mucha utilidad para la aplicación de tempranas y efectivas intervenciones preventivas o terapéuticas y los predictores permiten monitorizar la repuesta al tratamiento para reducir el riesgo de ECV. (Anexo 6, figura 15)

Un aspecto clave es que el M-IA predice el riesgo en solo dos categorías, el paciente se considera que está en riesgo de una enfermedad cardiovascular o, al contrario, no se considera en riesgo. Los M-Ts actuales estiman una probabilidad de riesgo y un elevado

número de pacientes lo clasifican como riesgo intermedio, con puntos de corte arbitrarios, sin acceso a otros estudios, no disponibles para una población tan amplia, en la atención primaria, que pudieran reclasificar el riesgo; de ahí que, no se realicen intervenciones efectivas en este grupo, que unido a la categoría riesgo bajo aporta el mayor número de episodios cardiovasculares.^{7,9,11} El M-IA al predecir el riesgo en solo dos categorías facilita una evaluación completa del riesgo, al mismo tiempo, ofrece acciones de promoción o prevención y control a toda la población.

Igualmente, el M-IA permite la detección de la inflamación de bajo grado, mediante la PCR, una prueba simple y accesible para un laboratorio de mediana complejidad y el autor sugiere su incorporación dentro de las pruebas de laboratorio solicitadas para evaluar el riesgo cardiovascular.^{158,159}

Una ventaja del M-IA implementado es que no solo clasifica los pacientes según el riesgo, sino que permite el diagnóstico de la dislipemia aterogénica, el síndrome metabólico, sospechar una aterosclerosis subclínica, infradiagnosticadas, infratratadas y, por tanto, infracontroladas en la atención primaria, de igual forma, utilizando la potencia del poder predictivo del riesgo, con la concentración de triglicéridos y el cHDL, en el cálculo del índice aterogénico del plasma TG/cHDL, que ofrece un valor único, junto a las otras condiciones clínicas que detecta el M-IA. Hay que destacar, es una guía para la acción en el control del riesgo en la atención primaria de salud.

En cuanto al valor diagnóstico de la glucemia alterada en ayunas, ante los criterios glucogénicos actuales que retrasan el diagnóstico de la DM2, pues a menudo los síntomas no son intensos o están ausentes, debido al ritmo lento con que avanza la hiperglucemia,^{148,160} el modelo predictivo propuesto permite sospechar resistencia a la

insulina y se puede prevenir o retrasar la DM2 y sus complicaciones que son progresivas y acumulativas según el tiempo de evolución.¹⁴⁹

El autor considera el equipo básico de salud en la comunidad, como el escenario ideal para intervenir, no solo con tratamiento farmacológico, sino también, cambiando estilos de vida no saludables y tratando de modificar las condiciones de vida para controlar el riesgo. Por otro lado, informar el nivel de riesgo podría permitir una motivación y adherencia al tratamiento por el paciente para modificarlo, previniendo o retrasando las ECV y también, al médico para ejercer una medicina integrativa¹⁶¹ individualizada, personalizada y de precisión y, por tanto, decidir cuándo y cómo brindar tratamiento, avanzando hacia un cuidado de salud integrativo, mejorando así la calidad de la atención médica.

Conviene especificar que el M-IA fue diseñado para predecir el riesgo en personas asintomáticas, es decir en prevención primaria, pero el autor considera, que se puede utilizar en prevención secundaria, pues la aplicación del modelo después de un episodio cardiovascular puede ayudar a identificar los pacientes con mayor riesgo de morbilidad y mortalidad,¹⁰⁵ asimismo, su utilización en edades avanzadas, restándole el gran peso de la edad como factor de riesgo en este grupo de personas y donde el objetivo sería evitar la ruptura de la placas ateroscleróticas inestables o de alto riesgo, abordando conjuntamente las características de las placas y la carga aterosclerótica total de los ancianos.^{162,163}

Otro elemento ya mencionado es que el modelo aporta objetivos o blancos para estrategias de intervención farmacológica orientadas a la regresión de las placas ateroscleróticas mediante la reducción de su contenido de macrófagos, lipídico y del estado inflamatorio. La progresión de la placa precede a su ruptura, varios tratamientos farmacológicos inducen la

regresión de la placa aterogénica, la evidencia confirma la relación entre la reducción del cLDL y la reducción de la placa; recientemente se puede determinar, además, su composición y volumen por técnicas de imagen invasivas o no.^{105,164} Resaltando el rol de la inflamación de bajo grado, el autor considera el uso de la pcr-as como blanco terapéutico, que puede ser proporcional a la reducción o el control de la placa aterogénica con la reducción de la inflamación y permite monitorizar o controlar la respuesta al tratamiento, más allá del cLDL, evitando así, el riesgo residual de origen lipídico.^{164,165}

En otro sentido, el M-IA implementado, aunque no fue comparado con otros M-Ts, solo con las tablas de la OMS 2019 y, por ende, según la evidencia que aporta la investigación y, basado en los métodos teóricos aplicados, vale señalar desde otra perspectiva, que es diferente y supera las limitaciones de los M-Ts.

En su opinión el M-IA moderno, que se presenta a la comunidad científica, cómo se ha mencionado, supera las limitaciones de las funciones de predicción actuales, y a diferencia de los M-Ts podría extrapolarse a la población cubana y a poblaciones foráneas, sin necesidad de recalibrar; así pues la investigación PESA³⁸ aporta evidencia de que aún recalibrados podrían ser inefectivos en la predicción fiable del riesgo, pues en las categorías bajo e intermedio se evidenció una alta incidencia de aterosclerosis subclínica con aumento de las concentraciones de pcr-as, triglicéridos, cHDL y cLDL normales o altas. Al igual que, los factores de riesgo tradicionales pueden variar en diferentes poblaciones ya que cada una tiene sus particularidades, no solo genéticas, sino también de hábitos nutricionales o estilos de vida y el M-IA incorpora predictores novedosos relacionados con la fisiopatología de la aterosclerosis, que el autor opina no deben variar en pacientes con

riesgo en diferentes poblaciones, esta idea requeriría una validación externa independiente del modelo predictivo.³

En la población cubana, utilizando este M-IA puede irse más allá que el propuesto por la iniciativa HEARTS¹⁴ para la predicción del riesgo cardiovascular,¹⁶⁶ que son las tablas de la OMS de 2019,¹ donde el M-IA permite identificar los pacientes con inflamación sistémica de bajo grado, un elemento clave en la aterosclerosis y que relaciona a varias condiciones y entidades, y a consideración, del autor con un diseño robusto con una predicción más efectiva que las tablas de la OMS 2019, validado en una fracción de la población cubana.¹¹³

Dentro de las limitaciones del M-IA puede señalarse en primer lugar que ningún modelo es perfecto, ya que son abstracciones de la realidad, sin embargo son útiles, pues en condiciones normales puede que los datos varíen por disimiles razones, además, no se pudo disponer de una base de datos de prueba por limitaciones logísticas, ni hacer comparaciones porque no se encontraron M-IA disponibles en la población cubana y está pendiente su validación externa independiente. Es preciso señalar que podrían utilizarse otros algoritmos más complejos con los mismos datos, pues la minería de datos es una ciencia experimental, existen una gran cantidad de técnicas de preprocesamiento, así como algoritmos de clasificación, pero no hay un método único, universal y mejor para un problema específico, se necesita encontrar los métodos de aprendizaje que trabajen mejor ante un determinado problema.^{15,113}

La tabla de la OMS es útil y se recomienda su uso en ausencia de otros modelos disponibles, los modelos predictivos son útiles: es mejor predecir que adivinar; la tabla de la OMS tiene limitaciones, ya que usan como predictor el colesterol total que no debe valorarse como una

cifra única, es importante, la proporción existente entre este y el cHDL, el cLDL y su valor respecto a los TG para una evaluación completa del perfil lipídico del paciente.¹⁶⁷

Otra limitación de las tablas de la OMS, al disponer de datos no analíticos, no se considera el colesterol y la diabetes y se usa como predictor el IMC, este mantiene su valor predictivo, pero deben determinarse también otras mediciones antropométricas para determinar la grasa visceral como: la circunferencia abdominal, el índice alométrico de forma corporal, el índice cintura/cadera y cadera/altura, por las limitaciones ya señaladas y el rol de la adiposidad visceral en la aterosclerosis^{140,146,167,168}

Evidencias recientes revelan que el IMC puede sobre o infraestimar la adiposidad y provee información inadecuada sobre la salud a nivel individual. Hoy, se define la obesidad clínica como una enfermedad crónica sistémica, causando daño en órganos y tejidos, que causan complicaciones cardiovasculares y renales, entre otras; y la obesidad preclínica como un estado de exceso de adiposidad, pero sin daños en órganos, pero que generalmente incrementa el riesgo de obesidad clínica; otras ENT como DM2, ECV, ciertos tipos de cáncer y desórdenes mentales; en síntesis, la preclínica contra la clínica, sería salud versus enfermedad. Se recomienda también, que el IMC debe usarse en población saludable, con fines de pesquiasaje, pero el exceso de adiposidad se debe confirmar con algún índice antropométrico de los mencionados, para medir la grasa corporal.^{169,170}

Un problema más amplio está relacionado con el exceso de adiposidad, la resistencia a la insulina y la metainflamación de bajo grado; de ahí la importancia del novedoso concepto de “salud metabólica”, que el M-IA diseñado con los nuevos predictores añadidos; a diferencia de los M-Ts, es capaz de identificar e incluir la promoción de salud a nivel individual y poblacional, para mejorar la salud metabólica, con beneficios más allá de la

salud cardiovascular, con un cambio de paradigma donde la prevención primaria debería ser la prevención de la aparición de la “disfunción metabólica”, en lugar de la prevención de la aparición de las enfermedades cardiovasculares y otras enfermedades incluidas la depresión y el cáncer. ^{168,171,172,173}

La prevención de la enfermedad cardiovascular basada en sus factores de riesgo es un proceso complejo y multifactorial. Sus intervenciones permiten el control de los factores de riesgo modificables y en la predicción del riesgo cardiovascular a medida que se analizan nuevos resultados, se irán perfeccionando los modelos predictivos. La IA ha liderado la revolución digital en el campo de la predicción de ECV, debido a la capacidad de procesamiento de datos más robusta, con menos restricciones de condición, pues se utilizan varios índices nuevos que consisten principalmente en algunos datos multimodo que los M-Ts no pueden reconocer ni utilizar en absoluto, sus modelos permiten la reintroducción de datos y su utilidad. Otra mejora interesante es que la selección de predictores podría ejecutarse automáticamente por IA en lugar del cálculo logarítmico clásico con mejor rendimiento, proporcionando así, una estrategia predictiva más prometedora que los modelos tradicionales. ^{3,153,154,168,}

El autor considera que las limitaciones de los M-Ts disponibles, es la brecha (*gap*, en inglés) por donde escapan pacientes sin una predicción correcta de su riesgo cardiovascular, que aportan considerablemente a la morbilidad y mortalidad cardiovascular. La hipótesis es el objeto de la verificación o confirmación, en forma de enunciado afirmativo, se llega al conocimiento por vía sintética y expresa que el investigador dispone de una respuesta anticipada al problema. En la investigación no se refuta, y acorde con los resultados, se acepta, ya que el M-IA implementado podría favorecer una predicción precisa del riesgo de

enfermedad cardiovascular, yendo más lejos con IA en una predicción efectiva del riesgo y ayudar a cerrar la brecha actual en la predicción y control del riesgo, lo que se traduciría en su impacto clínico relevante.

El modelo propuesto a la comunidad científica sería uno de los primeros en superar las limitaciones en las funciones de predicción del riesgo cardiovascular usadas en la población cubana, y nos lleva a la comprensión de la inteligencia artificial como una nueva forma de generar conocimiento al permitir enfoques más precisos en la investigación y la prevención de las enfermedades cardiovasculares en individuos y poblaciones.^{3,168,174,175,176,177}

3.4 Conclusiones del capítulo 3

Por los resultados alcanzados en la evaluación y la validación, los modelos predictivos son efectivos para respaldar su empleo en la predicción del riesgo de enfermedad cardiovascular en la atención primaria de salud.

El modelo predictivo implementado fue más eficaz que las tablas de la OMS, basándose en su diseño, la comparación de ambos y el salto cualitativo, que es un avance en el desarrollo pues el modelo IA supera al modelo de la OMS.

El modelo diseñado es una herramienta útil para la atención primaria ya que puede contribuir a una predicción eficaz del riesgo cardiovascular.

El M-IA es diferente y supera las limitaciones de las funciones de riesgo actuales y podría extrapolarse a otras regiones, ya que, los predictores añadidos al modelo están relacionados directamente con todas las etapas del proceso aterosclerótico.

Se valora la importancia del novedoso concepto de “salud metabólica”, yendo más lejos de los beneficios de la salud cardiovascular a la prevención de las disfunciones metabólicas a

nivel individual y poblacional, para ofrecer una mejor calidad de vida. La inteligencia artificial ha generado una revolución en la medicina y la salud como una nueva forma de generar conocimiento al permitir enfoques más precisos en la investigación y la prevención de las enfermedades cardiovasculares en individuos y poblaciones.

CONCLUSIONES

- ✓ El modelo predictivo diseñado resultó ser útil, válido, comprensible, con un bajo riesgo de sesgo y de aplicabilidad clínica, por tanto, se respalda su empleo en la atención primaria de salud en Cuba.
- ✓ El modelo de analítica predictiva implementado incluyó predictores de riesgo novedosos, relacionados estrechamente con la fisiopatología de la aterosclerosis, a diferencia de los modelos tradicionales y, por consiguiente, supera las limitaciones actuales en la predicción del riesgo de enfermedad cardiovascular.
- ✓ El modelo predictivo implementado realizó una predicción más efectiva del riesgo de enfermedad cardiovascular al compararlo con las tablas predictivas de la Organización Mundial de la Salud de 2019.

RECOMENDACIONES

- ✓ Implantar el modelo predictivo en la atención primaria de salud, como herramienta práctica para la predicción del riesgo cardiovascular, lo que favorecería la disminución de la morbilidad y mortalidad por enfermedades cardiovasculares.
- ✓ Implantar estrategias de promoción de salud a nivel individual y poblacional para mejorar la salud metabólica, más allá de la salud cardiovascular, en la atención primaria de salud.
- ✓ Realizar la validación externa del modelo predictivo en poblaciones foráneas.

REFERENCIAS BIBLIOGRÁFICAS

1. WHO CVD Risk Chart Working Group. World Health Organization cardiovascular disease risk charts: revised models to estimate risk in 21 global regions. Lancet Glob Health 2019 [Internet]. 2019 [citado 14 de enero de 2020]; 1-11. DOI: [http://dx.doi.org/10.1016/S2214-109X\(19\)30318-3](http://dx.doi.org/10.1016/S2214-109X(19)30318-3)
2. Vega Abascal J, Guimarães Mosqueda M, Vega Abascal L. Riesgo cardiovascular, una herramienta útil para la prevención de las enfermedades cardiovasculares. Rev Cubana Med Gen Integr [Internet]. 2011[citado 12 de diciembre de 2013]; 27(1):91-7. Disponible en: <http://scielo.sld.cu/pdf/mgi/v27n1/mgi10111.pdf>
3. Cai Y, Cai YQ, Tang LY, Wang YH, Gong M, Jing TC, et al. Artificial intelligence in the risk prediction models of cardiovascular disease and development of an independent validation screening tool: a systematic review. BMC Medicine [Internet]. 2024 [citado 21 de noviembre de 2024];22(1):56. DOI: <https://doi.org/10.1186/s12916-024-03273-7>
4. Anuario Estadístico de Salud. MINSAP. Cuba. [Internet]. 2012 [citado 26 de febrero de 2022]. Disponible en: http://files.sld.cu/dne/files/2013/04/anuario_2012.pdf
5. Anuario Estadístico de Salud. MINSAP. Cuba. [Internet]. 2020 [citado 26 de febrero de 2022]. Disponible en: <http://files.sld.cu/bvscuba/files/2021/08/Anuario-Estadistico-Español-2020-Definitivo.pdf>
6. D'Agostino RB, Pencina MJ, Massaro, JM, Coady S. Cardiovascular disease risk assessment: Insights from Framingham. Glob Heart [Internet]. 2013 [citado 7 de

octubre de 2014]; 8(1): 11–23. Disponible en:
<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3673738/>

7. Vega Abascal J, Guimar Mosqueda M R, Garcs Hernndez Y, Vega Abascal L, Rivas Estvez M. Prediccin de riesgo coronario y cardiovascular global en la atencin primaria de salud. CCM [Internet]. 2015 [citado 2 de marzo de 2017]; 19(2): 202-211. Disponible en: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1560-43812015000200003&lng=es
8. Pizarro R, Masson W. Estudios prospectivos poblacionales: las fortalezas de las cohortes histricas. Acta Gastroenterol Latinoam [Internet]. 2020 [citado 23 de marzo de 2021];5 0(4):382-7. Disponible en: <https://actagastro.org/estudios-prospectivos-poblacionales-las-fortalezas-de-las-cohortes-historicas/>
9. Castellano JM, Narula J, Castillo J, Fuster V. Promocin de la salud cardiovascular global: estrategias, retos y oportunidades. Rev Esp Cardiol. 2014 [citado 21 de marzo de 2018];67(9):724-30. Disponible en: <http://www.revespcardiol.org/es/vol-67-num-09/sumario/13019630/>
10. Shrivastava AK, Singh HV, Raizada A, Singh SK. C-reactive protein, inflammation and coronary heart disease. The Egyptian Heart Journal [Internet]. 2015 [citado 7 de octubre de 2020]; 67:89-97. Disponible en: <https://dx.doi.org/10.1016/j.ehj.2014.11.005>
11. Elosua R. Las funciones de riesgo cardiovascular: utilidades y limitaciones. Rev Esp Cardiol [Internet]. 2014 [citado 14 de enero de 2015];67(2):77-9. Disponible en: <https://www.revespcardiol.org/es/pdf/90267565/S300/>

12. Grillo-Martín M, Mezquia de Pedro N, González-García S. Tablas de predicción de riesgo cardiovascular empleadas en Cuba. Revista Cubana de Cardiología y Cirugía Cardiovascular [Internet]. 2023 [citado 20 de mayo de 2023];29(1):1408. Disponible en: <https://revcardiologia.sld.cu/index.php/revcardiologia/article/view/1408>
13. Organización Panamericana de la Salud. HEARTS en las Américas: Marco de evaluación para la mejora continua de la calidad en los centros de atención primaria de salud. Washington, D.C.: OPS; 2024. [Internet]. [citado 21 de noviembre de 2024]. Disponible en: <https://iris.paho.org/handle/10665.2/62372>
14. Organización Panamericana de la Salud. HEARTS: Paquete técnico para el manejo de las enfermedades cardiovasculares en la atención primaria de salud. Riesgo cardiovascular: Manejo de las ECV basado en el riesgo. Washington, D.C.: OPS; 2024. [Internet]. [citado 21 de noviembre de 2024]. DOI: <https://doi.org/10.37774/9789275328507>
15. Witten I, Frank E, Hall M. Data Mining: Practical Machine Learning Tools and Techniques. Tercera Ed. San Francisco, California: Morgan Kaufmman, Elsevier; 2011. p.525 Disponible en: https://www.academia.edu/23331284/Data_Mining_Practical_Machine_Learning_Tools_and_Techniques_3rd_Edition
16. Mazón B, Pinta M, Redovran F. Desarrollo de competencias en Minería de Datos, una experiencia didáctica. En: Sistematización de experiencias educativas innovadoras. Primera Edición. Machala, Ecuador: UTMACH; 2020. p. 460. Disponible en: https://www.researchgate.net/publication/343558044_Development_of_skills_in_Data_Mining_a_educational_experience

17. Vaquero de Miguel M. Aproximaciones a la Explicación de Decisiones Algorítmicas: Inteligencia Artificial Explicable [Internet] [Trabajo Fin de Grado]. [Madrid]: Universidad Politécnica de Madrid; 2020 [citado 14 de febrero de 2021]. Disponible en: <http://oa.upm.es/63381>
18. Bannister C, Halcox J, Currie C, Preece A, Spasic I. A genetic programming approach to development of clinical prediction models: A case study in symptomatic cardiovascular disease. PLoS ONE [Internet]. 2018 [citado 26 de febrero de 2022]. ;13(9): e0202685. DOI: <https://doi.org/10.1371/journal.pone.0202685>
19. Dimopoulos AC, Nikolaidou M, Caballero F, Engchuan W, Sanchez-Niubo A, Arndt H, et al. Machine learning methodologies versus cardiovascular risk scores, in predicting disease risk. BMC Medical Research Methodology [Internet]. 2018 [citado 14 de julio de 2020];18(179):2-11. DOI: <https://doi.org/10.1186/s12874-018-0644-1>
20. Kim, J.O., Jeong Y.S., Kim J.H., Lee JW. Park D., Kim H.S. Machine Learning-Based Cardiovascular Disease Prediction Model: A Cohort Study on the Korean National Health Insurance Service Health Screening Database. Diagnostics [Internet]. 2021 [citado 27 de julio de 2022];11(943). Disponible en: https://www.researchgate.net/publication/351854383_Machine_Learning-Based_Cardiovascular_Disease_Prediction_Model_A_Cohort_Study_on_the_Korean_National_Health_Insurance_Service_Health_Screening_Database
21. Anuario Estadístico de Salud. MINSAP. Cuba. [Internet]. 2023. Ed 2024 [citado 26 de noviembre de 2024]. Disponible en: <https://files.sld.cu/dne/files/2024/09/Anuario-Estadistico-de-Salud-2023-EDICION-2024.pdf>

22. Organización Mundial de la Salud. Enfermedades no transmisibles [Internet]. 2022 [citado 28 de abril de 2023]. Disponible en: <https://www.who.int/es/health-topics/noncommunicable-diseases>
23. Badawy M, Naing L, Johar S, Ong S, Rahman H, Tengah D, et al. Evaluation of cardiovascular diseases risk calculators for CVDs prevention and management: scoping review. BMC Public Health [Internet]. 2022 [citado 30 de abril de 2023];22(1):1742. DOI: <https://doi.org/10.1186/s12889-022-13944-w>
24. Organización Mundial de la Salud. Enfermedades cardiovasculares [Internet]. 2017 [citado 28 abr 2020]. Disponible en: [https://www.who.int/es/news-room/factsheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/es/news-room/factsheets/detail/cardiovascular-diseases-(cvds))
25. Fan J, Li X, Yu X, Liu Z, Jiang Y, Fang Y, et al. Global Burden, Risk Factors Analysis, and Prediction Study of Ischemic Stroke, 1990-2030. Neurology [Internet]. 2023 [citado 28 de mayo de 2023]; 17:10. DOI: <https://doi.10.1212/WNL.0000000000207387>
26. Organización Panamericana de la Salud. La carga de las enfermedades cardiovasculares en la Región de las Américas, 2000-2019. Portal de Datos de NMH [Internet]. 2021 [citado 28 de abril de 2023]. Disponible en: <https://www.paho.org/es/enlace/carga-enfermedades-cardiovasculares>
27. Organización Mundial de la Salud. Estimaciones de salud mundial 2019: defunciones por causa, edad, sexo, país y región, 2000-2019 [Internet]. 2020 [citado 28 de abril de 2023]. Disponible en: <https://www.who.int/es/news/item/09-12-2020-who-reveals-leading-causes-of-death-and-disability-worldwide-2000-2019>

28. CEPAL. Seminar on non-communicable diseases and their impact on sustainable development in the Caribbean [Internet]. 2021 [citado 27 de junio de 2022]. Disponible en: <https://www.cepal.org/es/node/54788>
29. Revueltas-Agüero M, Benítez-Martínez M, Hinojosa-Álvarez M del C, Venero-Fernández S, Molina-Esquivel E, Betancourt-Bethencourt JA. Caracterización de la mortalidad por enfermedades cardiovasculares: Cuba, 2009-2018. Archivo Médico Camagüey [Internet]. 2021 [citado 19 de septiembre de 2022];25(1):9-23. Disponible en: <https://revistaamc.sld.cu/index.php/amc/article/view/7707>
30. Anuario Estadístico de Salud. MINSAP 2018. Cuba. [Internet]. 2019 [citado 14 de diciembre de 2020]. Disponible en: <http://files.sld.cu/bvscuba/files/2019/04/Anuario-Electronico-Español-2018-ed-2019-compressed.pdf>
31. MINSAP. Prevención cardiovascular: alcanzable y muy necesaria. Cuba [Internet]. 2020 [citado 27 de octubre de 2022]. Disponible en: <https://salud.msp.gob.cu/prevencion-cardiovascular-alcanzable-y-muy-necesaria/>
32. Anuario Estadístico de Salud 2021. MINSAP. Cuba [Internet]. 2022 [citado 14 de octubre de 2022]. Disponible en: <https://files.sld.cu/dne/files/2022/10/Anuario-Estadistico-de-Salud-2021.-Ed-2022.pdf>
33. Anuario Estadístico de Salud 2014. MINSAP. Cuba [Internet] 2015 [citado 14 de octubre de 2015]. Disponible en: <http://files.sld.cu/bvscuba/files/2015/04/anuario-estadistico-de-salud-2014.pdf>
34. Anuario Estadístico de Salud 2016. MINSAP. Cuba. [Internet]. 2017 [citado 11 de julio de 2017]. Disponible en:

http://files.sld.cu/dne/files/2017/05/Anuario_Estadistico_de_Salud_ed_2016_edicion_2017.pdf

35. Fernández-Alvira JM, Fuster V, Pocock S, Sanz J, Fernández-Friera L, Laclaustra M, et al. Predicción de la aterosclerosis subclínica en individuos de bajo riesgo Puntuación de salud cardiovascular ideal y puntuación Fuster-BEWAT. JACC Ed Esp [Internet]. 2018 [citado 1 de julio de 2021];1(2):70-9. Disponible en: <http://jacc.elsevier.es/es-prediccion-aterosclerosis-subclnica-individuos-bajo-articulo-resumen-XX735109718626374>
36. Delgado L, Vázquez AM, Martínez-Sánchez G. Procesos moleculares patogénicos de la aterosclerosis y alternativas terapéuticas para su control. Rev Cubana Farm [Internet]. 2012 [citado 14 de enero de 2014];46(2). Disponible en: <http://scielo.sld.cu/pdf/far/v46n2/far14212.pdf>
37. Yaoming Y, Yang S, Xiaolin M. The role of fluid mechanics coronary atherosclerotic plaques: an up-to-date review. Rev Cardiovasc Med [Internet]. 2024 [citado 15 de abril de 2024];25(2): 49. DOI: <https://doi.org/10.31083/j.rcm2502049>
38. Echeverri D. ¿Somos conscientes de una nueva nomenclatura de aterosclerosis? Rev Colomb Cardiol [Internet]. 2017 [citado 12 de agosto de 2018]; 4:323-6. Disponible en: <http://www.scielo.org.co/pdf/rcca/v24n4/0120-5633-rcca-24-04-00323.pdf>
39. Castro M, Castro D, Seoane J, Torres L, González A, Aguirre C. Aterosclerosis subclínica en trabajadores de una institución de salud. Rev Cubana Med [Internet]. 2021 [citado 1 de abril de 2022]; 60(3): e2134. Disponible en: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0034-

40. Millán Núñez-Cortés J, Zamora A, Pintó X. Enfermedad vascular subclínica en pacientes con dislipemia aterogénica. Clin Investig Arterioscler [Internet]. 2017 [citado 1 de mayo de 2023]; 29:9-18. Disponible en: <https://www.elsevier.es/es-revista-clinica-e-investigacion-arteriosclerosis-15-articulo-enfermedad-vascular-subclinica-pacientes-con-X0214916817614222>
41. Fernández-Friera L, Peñalvo JL, Fernández-Ortiz A, Ibáñez B, López-Melgar B, Laclaustra M, et al. Prevalence, Vascular Distribution, and Multiterritorial Extent of Subclinical Atherosclerosis in a Middle-Aged Cohort: The PESA (Progression of Early Subclinical Atherosclerosis) Study. Circulation [Internet]. 2015 [citado 1 de mayo de 2023]; 131(24):2104-13. DOI: <https://doi/10.1161/CIRCULATIONAHA.114.014310>.
42. Núñez E, Fuster V, Gómez-Serrano M, Valdivielso J M, Fernández-Alvira J M, Martínez-López D, et al. Unbiased plasma proteomics discovery of biomarkers for improved detection of subclinical atherosclerosis. EBioMedicine [Internet]. 2022 [citado 14 de enero de 2023]; 76, 103874. DOI: <https://doi.org/10.1016/j.ebiom.2022.103874>
43. Hernández-Mijares A, Ascaso J, Blasco M, Brea A, Díaz A, Mantilla T, et al. Riesgo cardiovascular residual de origen lipídico. Componentes y aspectos fisiopatológicos. Clin Investig Arterioscler [Internet]. 2019 [citado 10 de julio de 2020]; 31 (2): 75-88. DOI: <https://doi.org/10.1016/j.arteri.2018.06.007>
44. Bialiou R. Dislipidemia aterogénica: Una bomba de tiempo en mi cuerpo [Internet]. 2022 [citado 5 de mayo de 2023]. Disponible en: <https://laboratoriosfarma.com/dislipidemia-aterogénica/>
45. Ascaso J, Millán J, Hernández-Mijares A, Blasco M, Brea A, Díaz A, et al. Dislipidemia aterogénica 2019. Documento de consenso del Grupo de Dislipidemia Aterogénica de

la Sociedad Española de Arteriosclerosis. Clin Investig Arterioscler [Internet]. 2020 [citado 3 de mayo de 2023]; 32 (3): 120-125. Disponible en: <https://www.elsevier.es/es-revista-clinica-e-investigacion-arteriosclerosis-15-articulo-dislipidemia-aterogena-2019-documento-consenso-S0214916820300024>

46. Pedro-Botet J, Ascaso J, Blasco M, Brea A, Díaz A, Hernández-Mijares A, et al. Triglicéridos, colesterol HDL y dislipemia aterogénica en la guía europea para el control de las dislipemias 2019 Clin Investig Arterioscler [Internet]. 2020 [citado 1 de mayo de 2023]; 32 (5): 209-218. Disponible en: <https://www.elsevier.es/es-revista-clinica-e-investigacion-arteriosclerosis-15-articulo-trigliceridos-colesterol-hdl-dislipidemia-aterogena-S0214916820300048>

47. Padrón Chacón R, Fraga Díaz C, Oliva Linares JE. El síndrome metabólico [Internet]. La Habana: Editorial Ciencias Médicas; 2019 [citado 7 de octubre de 2020]. 230 p. Disponible en: <http://www.ecimed.sld.cu/2019/09/04/el-sindrome-metabolico/>

48. Rodríguez Perón J, Rodríguez Izquierdo Mario. Síndrome metabólico y su correlación con ecuaciones de predicción del riesgo global de enfermedad cardiovascular. Rev Cubana Med Gen Integr [Internet]. 2022 [citado 5 de mayo de 2023]; 38(3): e1909. Disponible en: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0864-21252022000300016&lng=es

49. Ramírez Iñiguez MV, Vicente Herrero MT, López González AA, Capdevila García L. Síndrome metabólico y diabetes tipo 2. Estimación de riesgo en trabajadores aparentemente sanos. Medicina Balear [Internet]. 2020 [citado 22 de mayo de 2020];35(2):34-40. Disponible en:

https://www.researchgate.net/publication/342200330_Sindrome_metabolico_y_diabetes_tipo_2_Estimacion_de_riesgo_en_trabajadores_aparentemente_sanos

50. Vega Abascal J, Guimara Mosqueda M, Vega Abascal L. Mas alla de la definicion de sndrome metabolico: cido urico, protena c reactiva de alta sensibilidad y riesgo cardiovascular global. CCM [Internet]. 2017 [citado 7 de abril de 2018]; 21(1): 70-86. Disponible en: http://scieloprueba.sld.cu/scielo.php?script=sci_arttext&pid=S1560-43812017000100007&lng=es.
51. Mostaza JM, Pinto X, Armario P, Masana L, Ascaso JF, Valdivielso P, et al. Estndares SEA 2019 para el control global del riesgo cardiovascular. Clnica e Investigacion en Arteriosclerosis [Internet]. 2019 [citado 4 de mayo de 2023]; 31:1-43. Disponible en: <https://www.sciencedirect.com/science/article/pii/S0214916819300464>
52. Aray-Andrade M, Lopez-Montanero E, Preciado-Gomez J. Dislipidemia aterogenica, factor subdiagnosticado para las complicaciones cardiovasculares. Salud Pblica Mex [Internet]. 2018 [citado 7 de octubre de 2020]; 60(4): 381-382. Disponible en: http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0036-36342018000400007&lng=es
53. Ponte C, Isea- Perez J, Lorenzatti A, Lopez-Jaramillo P, Wyss F, Pinto X, et al. Dislipidemia aterogenica en Latino America: prevalencia, causas y tratamiento. Rev Venez Endocrinol Metab [Internet]. 2017 [citado 7 de octubre de 2020]; 15(2): 106-129. Disponible en: http://ve.scielo.org/scielo.php?script=sci_arttext&pid=S1690-31102017000200006&lng=es.
54. Lobos JM, Brotons C. Factores de riesgo cardiovascular y atencion primaria: evaluacion e intervencion. Aten Primaria [Internet]. 2011 [citado 7 de mayo de 2021];

43(12):668-677. Disponible en: <https://www.elsevier.es/es-revista-atencion-primaria-27-pdf-S0212656711004689>

55. Areiza M, Osorio E, Ceballos M, Amariles P. Conocimiento y factores de riesgo cardiovascular en pacientes ambulatorios. Rev Colomb Cardiol [Internet]. 2018 [citado 19 de abril de 2019]; 25(2):162-168. Disponible en: http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0120-56332018000200162&lng=en
56. Jaramillo PL, Félix C, Escudero C, López M. Factores de riesgo clásicos y emergentes para enfermedades cardiovasculares. MedUNAB [Internet]. 1998 [citado 17 de mayo de 2000];1(3):171-83. Disponible en: <https://revistas.unab.edu.co/index.php/medunab/article/view/389>
57. Brotons-Cuixart C, Fernández-Valverde D, Moral-Peláez I. Factores de riesgo cardiovascular emergentes frente a clásicos. Diabetes Práctica [Internet]. 2018 [citado 16 de enero de 2019]; 09(01):1-36. Disponible en: http://www.diabetespractica.com/files/1524476455.02_revision_dp-9-1.pdf
58. Quirós Fallas R. Actualización de los factores de riesgo cardiovascular. Revista Médica Sinergia [Internet]. 2017 [citado 7 de mayo de 2021]; 2(1):3-7. Disponible en: <https://www.medigraphic.com/pdfs/sinergia/rms-2017/rms171a.pdf>
59. Bensusán T. Factores de riesgo cardiovascular emergentes. 6° Curso de Capacitación de Posgrado a Distancia Síndrome Metabólico y Riesgo Vascular. Fepreva [Internet]. 2012 [citado 7 de octubre de 2017]. Disponible en: http://www.fepreva.org/curso/6to_curso/material/ut9.pdf

60. Stefan N, Schulze MB. Metabolic health and cardiometabolic risk clusters: implications for prediction, prevention, and treatment. *The Lancet Diabetes & Endocrinology* [Internet]. 2023 [citado 12 de octubre de 2023];0. Disponible en: [https://www.thelancet.com/journals/landia/article/PIIS2213-8587\(23\)00086-4/abstract](https://www.thelancet.com/journals/landia/article/PIIS2213-8587(23)00086-4/abstract)
61. Badimon L, Peña E, Arderiu G, Padró T, Slevin M, Vilahur G, et al. C-Reactive Protein in Atherothrombosis and Angiogenesis. *Front Immunol* [Internet]. 2018 [citado 1 de junio de 2020];9. Disponible en: <https://www.frontiersin.org/articles/10.3389/fimmu.2018.00430/full>
62. Martín-Ventura JL, Blanco-Colio LM, Tuñón J, Muñoz-García B, Madrigal-Matute J, Moreno J, et al. Biomarcadores en la medicina cardiovascular. *Rev Esp Cardiol* [Internet]. 2009 [citado 19 de abril de 2010]; 62(6):677-688. Disponible en: <https://www.revespcardiol.org/es-biomarcadores-medicina-cardiovascular-articulo-13137603>
63. Vega Abascal J, Piriz Assa A, Guimará Mosqueda M, Vega Abascal L, Caballero González L, Nápoles Riaño D. PredRCG_aps: Predicción del riesgo cardiovascular global en la atención primaria de salud. *CCM* [Internet]. 2022 [citado 19 de septiembre de 2022]; 26(2). Disponible en: <http://www.revcoemed.sld.cu/index.php/cocmed/article/view/4202/2136>
64. Álvarez A. Las tablas de riesgo cardiovascular. Una revisión crítica. *Medifam* [serie en internet]. 2001 [citado 27 de junio de 2009];11(3). Disponible en: http://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S113157682001000300002&lng=en&nrm=iso

65. Wilson PW, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. *Circulation* [Internet]. 1998 [citado 7 de octubre de 2000]; 97 (18): 1837-1847. Disponible en: <https://datascientest.com/es/inteligencia-artificial-definicion>
66. Almeida Galarza RO. Análisis de supervivencia aplicado a las pequeñas y medianas empresas constructoras en Ecuador en el período 2012-2017[Tesis]. [Quito, Ecuador]: Universidad Andina Simón Bolívar; 2019. 131p. Disponible en: <https://repositorio.uasb.edu.ec/bitstream/10644/8058/1/T3493-MGFARF-Almeida Analisis.pdf>
67. Conroy RM, Pyörälä K, Fitzgerald AP, Sans S, Menotti A, De Backer G, et al. Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. *Eur Heart J* [Internet]. 2003 [citado 7 de octubre de 2014]; 24 (11):987-1003. DOI: [https://doi10.1016/s0195-668x\(03\)00114-3](https://doi10.1016/s0195-668x(03)00114-3)
68. Rodríguez IM, Mendivelso FO. Diseño de cohorte. *Rev Médica Sanitas* [Internet] .2019 [citado 27 de junio de 2021];22(1):37-40. Disponible en: <https://revistas.unisanitas.edu.co/index.php/rms/article/view/432/349>
69. Molinero LM. Modelos de riesgo cardiovascular. Estudio Framingham. Proyecto SCORE [monografía en internet]. 2003 [citado 20 de diciembre de 2009]. Disponible en: <http://www.seh-lelha.org/modelries.htm>
70. Álvarez-Acevedo E. William Kannel y el estudio Framingham. *Revista Cubana de Medicina Militar* [Internet]. 2022 [citado 19 de mayo de 2023]; 51 (2) Disponible en: <https://revmedmilitar.sld.cu/index.php/mil/article/view/1732>

71. Pencina MJ, D'Agostino RB, Larson MG, Massaro JM, Vasan RS. Predicting the 30-year risk of cardiovascular disease: The Framingham heart study. *Circulation* [Internet] 2009 [citado 19 de mayo 2010]; 119 (24): 3078-3084. DOI: <https://doi.org/10.1161/CIRCULATIONAHA.108.816694>
72. Guisado Rasco A, Cristobo Sáinz P, Barón-Esquivias G. ¿Cuáles son las diferencias entre las funciones para la evaluación del riesgo cardiovascular? *Cardiocre* [Internet]. 2012 [citado 27 de junio de 2018];47(1):12-5. Disponible en: <https://www.elsevier.es/es-revista-cardiocre-298-articulo-cuales-son-diferencias-entre-funciones-S1889898X11001423>
73. Marrugat J, Solanas P, D'Agostino R, Sullivan L, Ordovas J, Cerdón F, et al. Estimación del riesgo coronario en España mediante la ecuación de Framingham calibrada. *Rev Esp Cardiol* [Internet]. 2003 [citado 20 de mayo de 2015];56(3):253-61. Disponible en: <http://www.revespcardiol.org/es-estimacion-del-riesgo-coronario-espana-articulo-13043951>
74. Olivera Solís R, González Rodríguez E. Modelos de estimación de riesgo cardiovascular. Las TIC y su aporte [Internet]. 2019 [citado 19 de enero de 2021]. Disponible en: <https://convencion.uclv.cu/es/event/xvi-simposio-internacional-de-pensamiento-latinoamericano-2019-3/track/modelos-de-estimacion-de-riesgo-cardiovascular-las-tic-y-su-aporte-2392>
75. Assmann G, Cullen P, Schulte H. Simple scoring scheme for calculating the risk of acute coronary events based on the 10-year follow-up of the prospective cardiovascular Münster (PROCAM) study. *Circulation* [Internet]. 2002 [citado 19 de enero de 2021];

105 (3): 310-315. Disponible en:
<https://www.ahajournals.org/doi/10.1161/hc0302.102575>

76. Hippisley-Cox J, Coupland C, Robson J, Brindle P. Derivation, validation, and evaluation of a new QRISK model to estimate lifetime risk of cardiovascular disease: cohort study using QRResearch database BMJ [Internet] 2010. [citado 7 de octubre de 2014]; 341: c6624. DOI://doi10.1136/bmj.c6624
77. World Health Organization. Prevention of cardiovascular disease: guidelines for assessment and management of cardiovascular risk [Internet]. Geneva: World Health Organization; 2007 [citado 11 de marzo de 2010]. Disponible en: <https://apps.who.int/iris/handle/10665/43685>
78. Varona--Pérez P, Armas-Rojas N, Suárez-Medina R, Bonet-Gorbea M, Dueñas-Herrera A. Estimación del riesgo cardiovascular en la población cubana. Una aproximación al tema. Rev Cubana Cardiol Cir Cardiovasc [Internet]. 2015 [citado 19 de mayo de 2018]; 21 (4) Disponible en: <https://revcardiologia.sld.cu/index.php/revcardiologia/article/view/606>
79. Acosta Batista C, Herrera Izquierdo G, Rivera Ledesma E, Mulling Pérez R, Martínez García R. Epidemiología de los factores de riesgo cardiovascular y riesgo cardiovascular global en personas de 40 a 79 años en atención primaria. CorSalud [Internet]. 2015 [citado 27 de junio de 2018];7(1):35-45. Disponible en: <http://www.corsalud.sld.cu/sumario/2015/v7n1a15/frc-rcv.html>
80. Peral Sánchez M, Alegret Rodríguez M, Guirado Cruz R. Estimación del riesgo cardiovascular en una población del área de salud del Policlínico Santa Clara. Medicentro Electrónica [Internet]. 2016 [citado 27 de octubre de 2018]; 20(1): 38-45.

Disponible en: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1029-30432016000100006&lng=es

81. Gaziano TA, Young CR, Fitzmaurice G, Atwood S, Gaziano JM. Laboratory-based versus non-laboratory-based method for assessment of cardiovascular disease risk: the NHANES I Follow-up Study cohort. *Lancet* [Internet]. 2008 [citado 20 de enero de 2015];371(9616):923-31. Disponible en: [https://doi10.1016/S0140-6736\(08\)60418-3](https://doi10.1016/S0140-6736(08)60418-3)
82. De la Noval García R, Romero Martínez S, Dueñas Herrera A, Armas Rojas N, Acosta González M, Ortega Torres Y. Estimación del Riesgo Cardiovascular Global en trabajadores del Instituto de Cardiología y Cirugía Cardiovascular. *Rev Cubana Cardiol Cir Cardiovasc* [Internet]. 2013 [citado 20 de abril de 2019];19(2):66-71. Disponible en: <https://revcardiologia.sld.cu/index.php/revcardiologia/article/view/355>
83. Santoyo Rodríguez FA, Hernández Hernández Z, Hecheverría Nassar L, Meireles Delgado DM, Rojas Iriarte C. Riesgo cardiovascular global según tablas de Gaziano en pacientes hipertensos. *Panorama Cuba y Salud* [Internet]. 2018 [citado: 13 de diciembre de 2020];13(1). Disponible en: http://www.revpanorama.sld.cu/index.php/panorama/article/view/771/pdf_135
84. Revueltas-Agüero M, Valdés-González Y, Serra-Larín S, Suárez-Medina R, Ramírez-Sotolongo JC. Estimación del riesgo cardiovascular en una población, según dos tablas predictivas. *Arch Méd Camagüey* [Internet]. 2020 [citado 7 de febrero de 2021];24(5). Disponible en: <http://www.revistaamc.sld.cu/index.php/amc/article/view/7533/3636>
85. Armas Rojas N, Dueñas Herrera A, Suárez Medina R, Llerena Rojas L, de la Noval García R, Varona Pérez P, et al. Estimación del Riesgo Cardiovascular Global en el Municipio Colon. Matanzas, Cuba. *Rev Cubana Cardiol Cir Cardiovasc* [Internet]. 2016

[citado: 26 de abril de 2020];22(3). Disponible en:
http://www.revcardiologia.sld.cu/index.php/revcardiologia/article/view/656/html_51

86. Inteligencia artificial: definición, historia, usos, peligros. Formation Data Science [Internet]. 2022 [citado 5 de junio de 2023]. Disponible en:
<https://datascientest.com/es/inteligencia-artificial-definicion>

87. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, Depristo M, Chou K, et al. A guide to deep learning in health care. Nat Med [Internet]. 2019 [citado 19 de septiembre de 2021];25(1): 24-29. Disponible en:
https://www.researchgate.net/publication/330203264_A_guide_to_deep_learning_in_healthcare

88. Charfe F. Cómo es el proceso de extraer conocimiento a partir de bases de datos | campus MVP.es [Internet]. 2020 [citado 27 de junio de 2023]. Disponible en:
<https://www.campusmvp.es/recursos/post/el-proceso-de-extraccion-de-conocimiento-a-partir-de-bases-de-datos.aspx>

89. Gironés J, Casas J, Minguillón J, Caihuelas R. Minería de datos: modelos y algoritmos [Internet]. Primera Ed. Barcelona, España: UOC; 2017 [citado 27 de junio de 2020]. 273 p. Disponible en: <https://dokumen.site/file/mineria-de-datos-modelos-y-algoritmospdf-a5b39f02c4ae11>

90. Piatetsky-Shapiro G, Frawley W J. Knowledge Discovery in Databases [Internet] AAAI/MIT Press; 1991 [citado 7 de octubre de 2020]. 540 p. Disponible en:
<https://mitpress.mit.edu/9780262660709/knowledge-discovery-in-databases/>

91. Saleh Basma Jumaa, Saedi Ahmed Yousif Falih, Al-Aqbi, Ali Talib Qasim, Salman Lamees Abdalhasan . A review paper: analysis of weka data mining techniques for heart disease prediction system. Library Philosophy and Practice (e-journal). [Internet]. 2020 [citado 20 de noviembre de 2024]; 4032. Disponible en: <https://digitalcommons.unl.edu/libphilprac/4032>
92. Bases de Datos KDD ¿Qué Son? [Internet]. Salud Electrónica. 2021 [citado 10 de julio de 2023]. Disponible en: <https://saludelectronica.com/descubrimiento-de-conocimiento-en-bases-de-datos-kdd/>
93. Aman Chillar RS. Analyzing Predictive Algorithms in Data Mining for Cardiovascular Disease using WEKA Tool. International Journal of Advanced Computer Science and Applications (IJACSA) [Internet]. 2021 [citado 7 de diciembre de 2024];12(8). Disponible en: <https://thesai.org/Publications/ViewPaper?Volume=12&Issue=8&Code=IJACSA&SerialNo=17>
94. Gallego Valcárcel DA, Lucas Monsalve DF. Modelos de aprendizaje automático para la predicción del riesgo de fatalidad por insuficiencia cardíaca con datos clínicos [Internet] [Trabajo de grado para optar al título de Ingeniero de Sistemas y Computación]. [Bogotá, Colombia]: Universidad Antonio Nariño Facultad de Ingeniería de Sistemas; 2021 [citado 27 de junio de 2022]. Disponible en: <http://repositorio.uan.edu.co/bitstream/123456789/4803/5/2021DavidAlejandroGallegoMonografia.pdf>

95. Espinosa Mina MA. Weka, áreas de aplicación y sus algoritmos: una revisión sistemática de literatura. *Ecociencia* [Internet]. 2018 [citado 6 de febrero de 2020]; 5:1-26. Disponible en: <https://revistas.ecotec.edu.ec/index.php/ecociencia/article/view/153>
96. Espino Timón C. Análisis predictivo: técnicas y modelos utilizados y aplicaciones del mismo - herramientas Open Source que permiten su uso [Internet] [Tesis de Grado en Ingeniería Informática]. Universidad de Cataluña; 2017. Disponible en: <https://core.ac.uk/display/77515842>
97. Centeno A. Big Data. Técnicas de machine learning para la creación de modelos predictivos para empresas [Internet] [Trabajo fin de carrera]. Universidad Pontificia de Comillas; 2020 [citado 4 de junio de 2021]. Disponible en: [https://repositorio.comillas.edu/xmlui/bitstream/handle/11531/45878/Tecnicas demachinelearningparalacreaciondemodelospredictivosparaempresas_Centeno_Martin-Romero_Alfonso.pdf](https://repositorio.comillas.edu/xmlui/bitstream/handle/11531/45878/Tecnicas%20demachinelearningparalacreaciondemodelospredictivosparaempresas_Centeno_Martin-Romero_Alfonso.pdf)
98. Weka 3.9.5 Disponible en: <http://www.cs.waikato.ac.nz/ml/weka/>
99. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P. SMOTE: Synthetic Minority Over-sampling Technique. *Artificial Intelligence Research* [Internet]. 2002 [citado 27 de julio de 2003];341-378. Disponible en: <https://www.jair.org/index.php/jair/article/view/10302>
100. Moons KGM, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies: Explanation and Elaboration. *Ann Intern Med*. [Internet]. 2019 [citado 20 de noviembre de 2024];170(1): W1-W33. DOI: 10.7326/M18-1377

101. Fernandez-Felix B, Lopez- Alcalde J, Roque M, Muriel A, Zamora J. CHARMS and PROBAST at your fingertips: a template for data extraction and risk of bias assessment in systematic reviews of predictive models. BMC Medical Research Methodology 2023 [Internet]. [citado 23 de noviembre de 2024]. Disponible en: <https://bmcmmedresmethodol.biomedcentral.com/articles/10.1186/s12874-023-01849-0>
102. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Calster BV, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. The BMJ [Internet]. 2024 [citado 20 de noviembre de 2024];385: e078378. Disponible en: <https://pmc.ncbi.nlm.nih.gov/articles/PMC11019967/>
103. World Medical Association. World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects. JAMA [Internet]. 2013 [citado 7 de enero de 2015];310(20):2191-4. DOI: <https://doi.org/10.1001/jama.2013>.
104. Flores M, Barquera S, Carrión C, Rojas R, Villalpando S, Olaiz-Fernández G, et al. Concentraciones de proteína C reactiva en adultos mexicanos: alta prevalencia de un factor de riesgo cardiovascular. Salud Pública Méx [revista en la Internet]. 2007 [citado 12 de diciembre de 2024]; 49(Supl 3): s348-s360. Disponible en: http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0036-363420070009000006&lng=es.
105. Sáenz-San Martín A, Méndez-Ocampo P, Gutiérrez-Moctezuma I, Amezcua-Guerra LM. Proteína C reactiva, aspectos cardiovasculares de una proteína de fase aguda: una actualización para el médico. Arch Cardiol Mex [Internet]. 2024 [citado 12 de

- diciembre de 2024];94(2):191-202. Disponible en:
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11160552/>
106. Armas Rojas N, Lacey B, Soni M, Charles S, Carter J, Varona Pérez P, et al. Body-mass index, blood pressure, diabetes and cardiovascular mortality in Cuba: prospective study of 146,556 participants. BMC Public Health [Internet]. 2021 [citado 9 de octubre de 2022];21(963):2-10. DOI: <https://doi.org/10.1186/s12889-021-10911-9>
107. Colectivo de autores. Enfermedades no trasmisibles en Cuba. [Internet]. Tamayo Muñiz S, Pérez Perea L, Pérez González RD, coordinadores. La Habana: Editorial Ciencias Médicas; 2022. [citado 18 de marzo de 2023]. Disponible en: <http://www.bvscuba.sld.cu/libro/enfermedades-notransmisibles-en-cuba>
108. Chávez- Bósquez Ó, Torres-Vásquez M, Hernández -Torruco J, Hernández- Ocaña B. Impacto de los algoritmos de sobremuestreo en la clasificación de subtipos principales del síndrome de Guillain-Barré. Ingenius: Revista de Ciencia y Tecnología [Internet]. 2021 [citado 12 de diciembre de 2024];(25):20-31. Disponible en: <https://dialnet.unirioja.es/servlet/articulo?codigo=7773114>
109. Espinar Lara R. Modelos de Clasificación con datos no balanceados [Internet] [Trabajo Fin de Grado]. [España]: Universidad de Sevilla; 2018. [citado 27 de junio de 2020]. Disponible en: <https://idus.us.es/handle/11441/77518>
110. Chávez Cárdenas M C. Mejoras en la clasificación de interacciones de proteínas de secuencias de la Arabidopsis Thaliana utilizando técnicas de bases de datos desbalanceadas. Revista Cubana de Ciencias Informáticas [Internet]. 2019 [citado 27 de junio de 2023];13(3):91-106. Disponible en: http://scielo.sld.cu/scielo.php?script=sci_abstract&pid=S2227-18992019000300091&lng=es&nrm=iso&tlng=es

111. Elreedy D, Atiya AF, Kamalov F. A theoretical distribution analysis of synthetic minority oversampling technique (SMOTE) for imbalanced learning. Mach Learn [Internet]. 2023 [citado 27 de diciembre de 2023]; DOI: <https://doi.org/10.1007/s10994-022-06296-4>
112. Joloudari JH, Marefat A, Nematollahi MA, Oyelere SS, Hussain S. Effective Class-Imbalance learning based on SMOTE and Convolutional Neural Networks. arXiv [Internet]. 2022 [citado 17 de enero de 2023]. Disponible en: <http://arxiv.org/abs/2209.00653>
113. Vega-Abascal J, Piriz-Assa A, Nápoles-Riaño D. Modelo predictivo de enfermedad cardiovascular basado en inteligencia artificial en la atención primaria de salud. Revista Cubana de Medicina General Integral [Internet]. 2023 [citado 23 de octubre de 2023]; 39 (3): e2768. Disponible en: <https://revmgi.sld.cu/index.php/mgi/article/view/2768>
114. Ayala E, López R, Menéndez V. Modelos predictivos de riesgo académico en carreras de computación con minería de datos educativos. RED. [Internet]. 2021 [citado 14 de octubre de 2021];21(26). DOI: <https://doi.org/10.6018/red.463561>
115. Arnejo Calviño HA. Métodos para la mejora de predicciones en clases desbalanceadas en el estudio de bajas de clientes (CHURN) [Internet] [Master en Técnicas Estadísticas]. [La Coruña. España]: Universidad de Vigo; 2016 [citado 20 de agosto de 2021]. Disponible en: http://eio.usc.es/pub/mte/descargas/proyectosfinmaster/proyecto_1469.pdf
116. Abuhaija B, Alloubani A, Almatari M, Jaradat GM, Abdallah HB, Abualkishik AM, et al. A comprehensive study of machine learning for predicting cardiovascular disease using Weka and SPSS tools. International Journal of Electrical and Computer

- Engineering (IJECE) [Internet]. 2023 [citado 13 de diciembre de 2024];13(2):1891-902.
Disponible en: <https://ijece.iaescore.com/index.php/IJECE/article/view/29020>
117. Jain A, Tiwari S, Madhusudhana Subramanyam M, Jamal K, Yadav AS, Kumar S. Comparative study of Jrip j48 and naive bayes algorithm in Flower specie prediction. J Phys: Conf Ser [Internet]. 2021 [citado 12 de diciembre de 2025];1854(1):012046. Disponible en: <https://dx.doi.org/10.1088/1742-6596/1854/1/012046>
118. Worden K, Tsialiamanis G, Cross EJ, Rogers TJ. Artificial Neural Networks. En: Rabczuk T, Bathe KJ, editores. Machine Learning in Modeling and Simulation: Methods and Applications [Internet]. Cham: Springer International Publishing; 2023 [citado 20 de diciembre de 2024]. p. 85-119. DOI: https://doi.org/10.1007/978-3-031-36644-4_2
119. Kufel J, Bargieł-Łaczek K, Kocot S, Kozlik M, Bartnikowska W, Janik M, et al. What Is Machine Learning, Artificial Neural Networks and Deep Learning? -Examples of Practical Applications in Medicine. *Diagnostics* [Internet]. 2023 [citado 20 de diciembre de 2024];13(15): 2582. DOI: <https://doi.org/10.3390/diagnostics13152582>
120. Barredo Arrieta A, Díaz – Rodríguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, et al. “Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI”, *Information Fusion*, vol. 58, pp. 82-115, [Internet] 2020. [citado 27 de junio de 2021]. DOI: <https://doi.org/10.1016/j.inffus.2019.12.012>
121. Altwaijry N. Probability-Based Synthetic Minority Oversampling Technique. *IEEE Access* [Internet]. 2023. [citado 7 de octubre de 2023];(11): 28831-39. DOI: <https://doi.10.1109/ACCESS.2023.3260723>
122. Abella Miravet B. Mejora de las predicciones en muestras desbalanceadas [Internet] [Trabajo de fin de Grado en Ingeniería Informática]. [España]: Universidad Autónoma

- de Madrid; 2021 [citado 10 de julio de 2022]. Disponible en: <https://repositorio.uam.es/handle/10486/697900>
123. Szeghalmy S, Fazekas A. A Comparative Study of the Use of Stratified Cross-Validation and Distribution-Balanced Stratified Cross-Validation in Imbalanced Learning. *Sensors* [Internet]. 2023 [citado 27 de junio de 2023];23(2333). Disponible en: <https://www.mdpi.com/1424-8220/23/4/2333>
124. Rosemberg A, Street A, Valladão DM, Hentenyck PV. Two-Stage ML-Guided Decision Rules for Sequential Decision Making under Uncertainty [Internet]. arXiv; 2024 [citado 20 de diciembre de 2024]. Disponible en: <http://arxiv.org/abs/2405.14973>
125. Badawy M, Naing L, Johar S, Ong S, Rahman HA, Tengah D, et al. Evaluation of cardiovascular diseases risk calculators for CVDs prevention and management: scoping review. *BMC Public Health* [Internet]. 2022 [citado 5 de abril de 2023];22(1):1742. DOI: <https://doi.org/10.1186/s12889-022-13944-w>
126. Saragossi J. Research & Subject Guides: Evidence-Based Medicine: Levels of Evidence [Internet]. 2023 [citado 9 de octubre de 2023]. Disponible en: https://guides.library.stonybrook.edu/evidence-based-medicine/levels_of_evidence
127. Hosein A, Stoute V, Chadee S, Singh NR. Evaluating Cardiovascular Disease (CVD) risk scores for participants with known CVD and non-CVD in a multiracial/ethnic Caribbean sample. *Peer J* [Internet]. 2020 [citado 10 de enero de 2023];8: e8232. DOI: <https://doi.org/10.7717/peerj>
128. Studziński K, Tomasiak T, Krzysztoń J, Józwiak J, Windak A. Effect of using cardiovascular risk scoring in routine risk assessment in primary prevention of cardiovascular disease: an overview of systematic reviews. *BMC Cardiovascular*

- Disorders [Internet]. 2019 [citado 8 de julio de 2021];19(1):11. DOI: <https://doi.org/10.1186/s12872-018-0990-2>
129. Babic N, Valjevac A, Zaciragic A, Avdagic N, Zukic S, Hasic S. The Triglyceride/HDL Ratio and Triglyceride Glucose Index as Predictors of Glycemic Control in Patients with Diabetes Mellitus Type 2. *Med Arch* [Internet]. 2019 [citado 10 de julio de 2021];73(3):163-8. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6643328/>
130. Rodríguez-Martínez BS, Hernández-González MA, Borrayo-Sánchez G, Solorio-Meza SE, Medina-Torres ÓM, Gutiérrez-Galván MX, et al. Índice aterogénico y su relación con la mortalidad del síndrome isquémico coronario agudo *Rev Med Inst Mex Seguro Soc.* [Internet]. 2023 [citado 10 de enero de 2024] 18;61(Suppl 2): S226-S232. Disponible en: <https://pubmed.ncbi.nlm.nih.gov/38016097/>
131. Ouchi G, Komiya I, Taira S, Wakugami T, Ohya Y. Triglyceride/low-density-lipoprotein cholesterol ratio is the most valuable predictor for increased small, dense LDL in type 2 diabetes patients. *Lipids in Health and Disease* [Internet]. 2022 [citado 7 de julio de 2023];21(1):4. DOI: <https://doi.org/10.1186/s12944-021-01612-8>
132. Li YW, Kao TW, Chang PK, Chen WL, Wu LW. Atherogenic index of plasma as predictors for metabolic syndrome, hypertension and diabetes mellitus in Taiwan citizens: a 9-year longitudinal study. *Sci Rep* [Internet]. 2021 [citado 20 de diciembre de 2024];11(1):9900. Disponible en: <https://www.nature.com/articles/s41598-021-89307-z>
133. Kosmas C, Rodríguez S, Bousvarou M, Papakonstantinou E, Peña E, Guzman E, et al. The Triglyceride/High-Density Lipoprotein Cholesterol (TG/HDL-C) Ratio as a Risk Marker for Metabolic Syndrome and Cardiovascular Disease. *Diagnostics* [Internet].

- 2023 [citado 20 de diciembre de 2024];13, 929.DOI:
<https://doi.org/10.3390/diagnostics13050929>
134. Chen X, Zhang Z, Qiao G, Sun Z, Lu W. Immune and inflammatory insights in atherosclerosis: development of a risk prediction model through single-cell and bulk transcriptomic analyses. *Front Immunol* [Internet]. 2024 [citado 20 de diciembre de 2024];15. Disponible en:
<https://www.frontiersin.org/journals/immunology/articles/10.3389/fimmu.2024.1448662/full>
135. Rojas N, Caballero G, Madariaga A. Factores de riesgo tradicionales de enfermedad cardiovascular en el adulto mayor y su valor predictivo. *Revista Cubana de Medicina* [Internet]. 2023 [citado 20 de julio de 2023];62(1). Disponible en:
<https://revmedicina.sld.cu/index.php/med/article/view/2982>
136. Wong JJ, Hong R, Teo LLY, Tan RS, Koh AS. Atherosclerotic cardiovascular disease in aging and the role of advanced cardiovascular imaging. *Cardiovasc Health* [Internet]. 2024 [citado 21 de diciembre de 2024];1(1):1-8. Disponible en:
<https://www.nature.com/articles/s44325-024-00012-y>
137. Engels F. Anti-Dühring. Capítulo XI. Moral y derecho. libertad y necesidad [Internet]. 2016 [citado 25 de julio de 2022]. Disponible en: <https://inter-rev.foroactivo.com/t6105-f-engels-anti-duhring-capitulo-xi-moral-y-derecho-libertad-y-necesidad>
138. Reyes-Calderón JR. Karl Marx, dialéctica material de la historia. *Hallazgos* [Internet]. 2020 [citado 25 de julio de 2023];17(33):163-96. Disponible en:
http://www.scielo.org.co/scielo.php?script=sci_abstract&pid=S1794-38412020000100163&lng=en&nrm=iso&tlng=es

139. Rodríguez Acosta J, Hernández Rodríguez J, Cabrera Rode E, Cubas Dueñas I. Generalidades sobre el comportamiento de la proteína C reactiva en personas obesas. Rev Cubana Endocrinol [Internet]. 2023 [citado 24 de diciembre de 2024];34(1). Disponible en: <https://revendocrinologia.sld.cu/index.php/endocrinologia/article/view/376>
140. González-Gil EM, Peruchet-Noray L, Sedlmeier AM, Christakoudi S, Biessy C, Navionis AS, et al. Association of body shape phenotypes and body fat distribution indexes with inflammatory biomarkers in the European Prospective Investigation into Cancer and Nutrition (EPIC) and UK Biobank. BMC Medicine [Internet]. 2024 [citado 4 de diciembre de 2024];22(1):334. Disponible en: <https://doi.org/10.1186/s12916-024-03544-3>
141. Ridker PM, Buring JE, Rifai N, Cook NR. Development and Validation of Improved Algorithms for the Assessment of Global Cardiovascular Risk in Women: The Reynolds Risk Score. JAMA. [Internet]. 2007 [citado 4 de enero de 2020];297(6):611–619. DOI:10.1001/jama.297.6.611
142. Chunyu L, Chihua L. C-reactive protein and cardiovascular diseases: a synthesis of studies based on different designs. European Journal of Preventive Cardiology [Internet]. 2023 [citado 25 de diciembre de 2024];30(15): 1593-1596. DOI: <https://doi.org/10.1093/eurjpc/zwad116>
143. Han E, Fritzer-Szekeres M, Szekeres T, Gehrig T, Gyöngyösi M, Bergler-Kleina J. Comparison of High-Sensitivity C-Reactive Protein vs C-reactive Protein for Cardiovascular Risk Prediction in Chronic Cardiac Disease. The Journal of Applied Laboratory Medicine [Internet]. 2022 [citado 25 de diciembre de 2024]. Disponible en: <https://academic.oup.com/jalm/article/7/6/1259/6711153>

144. Kuppa A, Tripathi H, Al-Darraj A, Tarhuni W, Abdel-Latif A. Reactive Protein Levels and Risk of Cardiovascular Diseases: A Two-Sample Bidirectional Mendelian Randomization Study. *Int. J. Mol. Sci.* [Internet]. 2023 [citado 25 de diciembre de 2024]; 24,9129. DOI: 9129. <https://doi.org/10.3390/ijms24119129>
145. Fu Y, Wu Y, Liu E. C reactive protein and cardiovascular disease: From animal studies to the clinic (Review). *Experimental and Therapeutic Medicine* [Internet]. 2020 [citado 25 de diciembre de 2024]; 20, 1211-1219. DOI: <https://doi.org/10.3892/etm.2020.8840>
146. Vega Abascal J, Leyva Sicilia Y, Teruel Ginés R. La circunferencia abdominal. Su inestimable valor en la Atención Primaria de Salud CCM [Internet]. 2019 [citado 7 de abril de 2019];23(1):270-274.Disponible en: <http://www.revcocmed.sld.cu/index.php/cocmed/article/view/3023/1369>
147. Vera-Ponce VJ, Rodas-Alvarado L, Talavera JE, Cruz-Ausejo L, Torres-Malca JR, et al. Asociación entre resistencia a la insulina y proteína C reactiva en una muestra de peruanos no obesos. *Revista del Cuerpo Médico Hospital Nacional Almanzor Aguinaga Asenjo* [Internet]. 2021 [citado 16 de julio de 2022];14(2):124-7. Disponible en: http://www.scielo.org.pe/scielo.php?script=sci_abstract&pid=S2227-47312021000200002&lng=es&nrm=iso&tlng=es
148. Santos Lozano E. Resistencia a Insulina: Revisión de literatura. *Rev Méd Hondur* [Internet]. 2022 [citado 20 de julio de 2023]; 90 (1): 63-70. DOI: <https://doi.org/10.5377/rmh.v90i1.13824>
149. Adeva M, Martínez-Rodríguez J, González-Lucán M, Fernández-Fernández C, Castro-Quintela E. Insulin resistance is a cardiovascular risk factor in humans. *Diabetes*

- & Metabolic Syndrome: Clinical Research & Reviews [Internet]. 2019 [citado 10 de julio de 2020];13(2):1449-55. DOI: <https://doi.org/10.1016/j.dsx.2019.02.023>
150. Raposeiras-Roubin Sergio, Rosselló Xavier, Oliva Belén, Fernández-Friera Leticia, Mendiguren José M., Andrés Vicente, et al. Triglycerides and Residual Atherosclerotic Risk. Journal of the American College of Cardiology [Internet]. 2021 [citado 27 de junio de 2022];77(24):3031-41. Disponible en: <https://www.jacc.org/doi/10.1016/j.jacc.2021.04.059>
151. Sánchez-Cabo F, Rossello X, Fuster V, Benito F, Manzano JP, Silla JC, et al. Machine Learning Improves Cardiovascular Risk Definition for Young, Asymptomatic Individuals. JACC [Internet]. 2020 [citado 27 de junio de 2022]; 76 (14) 1674–1685.DOI: <https://doi.org/10.1016/j.jacc.2020.08.017>
152. Jiang Y, Zhang X, Ma R, Wang X, Liu J, Keerman M, et al. Cardiovascular Disease Prediction by Machine Learning Algorithms Based on Cytokines in Kazakhs of China. Clin Epidemiol [Internet]. 2021 [citado 27 de junio de 2024]; 13:417-28. DOI: 10.2147/CLEP.S313343
153. Ghiasi Hafezi S, Sahranavard T, Kooshki A, Hosseini M, Mansoori A, Fakhrian EA, et al. Predicting high sensitivity C-reactive protein levels and their associations in a large population using decision tree and linear regression. Sci Rep [Internet]. 2024 [citado 31 de diciembre de 2024];14(1):30298. Disponible en: <https://www.nature.com/articles/s41598-024-81714-2>
154. Cheng CH, Lee BJ, Nfor ON, Hsiao CH, Huang YC, Liaw YP. Using machine learning-based algorithms to construct cardiovascular risk prediction models for Taiwanese adults based on traditional and novel risk factors. BMC Med Inform Decis Mak [Internet].

- 2024 [citado 31 de diciembre de 2024]; 24:199. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11265113/>
155. Fonseca FAH, de Oliveira Izar MC. High-Sensitivity C-Reactive Protein and Cardiovascular Disease Across Countries and Ethnicities. *Clinics (Sao Paulo)* [Internet]. 2015 [citado 25 de diciembre de 2024];71(4):235-42. Disponible en: <http://www.elsevier.es/es-revista-clinics-22-articulo-high-sensitivity-c-reactive-protein-cardiovascular-disease-S1807593222011152>
156. Ridker P, Vinayaga M, Cook N, Rifai N, Lee I, Buring J. Inflammation, Cholesterol, Lipoprotein(a), and 30-Year Cardiovascular Outcomes in Women. *New England Journal of Medicine* [Internet]. 2024 [citado 31 de diciembre de 2024]; 391: 2087-97. Disponible en: <https://www.nejm.org/doi/full/10.1056/NEJMoa2405182>
157. Di Fabio JL, Gofin R, Gofin J. Análisis del sistema de salud cubano y del modelo atención primaria orientada a la comunidad. *Rev Cub Salud Pública* [Internet]. 2020 [citado 2 de agosto de 2023];46: e2193. Disponible en: <https://scielosp.org/article/rcsp/2020.v46n2/e2193/es>
158. Fiallos J, Martínez E. Proteína C reactiva como biomarcador de procesos inflamatorios [Internet] [Trabajo de Titulación para optar al título de Licenciada en Ciencias de la Salud en Laboratorio Clínico e Histopatológico]. [Ecuador]: Universidad Nacional de Chimborazo; 2022. Disponible en: <http://dspace.unach.edu.ec/handle/51000/9496>
159. Fernández-Mestre M, Salazar-Alcalá E, Matos-González G, Márquez I. Polimorfismo de genes de citocinas: ¿factores de riesgo cardiovascular en la población venezolana? *Arch Cardiol Mex* [Internet]. 2021 [citado 7 de octubre de 2022];91(3):281-8. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8351660/>

160. Pérez F, Rodrigo J, Salinas C. Limitaciones de los criterios de diagnóstico de la diabetes tipo 2 y la intolerancia a la glucosa. *Revista de Investigación Clínica* [Internet]. 2000 [citado 26 de noviembre de 2005];52(2):177-84. Disponible en: <https://dialnet.unirioja.es/servlet/articulo?codigo=4383735>
161. Grace S, Higgs J. Integrative medicine: enhancing quality in primary health care. *J Altern Complement Med* [Internet]. 2010 [citado 2 de agosto de 2023];16(9):945-50. Disponible en: <https://pubmed.ncbi.nlm.nih.gov/20809810/>
162. Tyrrell DJ, Goldstein DR. Ageing and atherosclerosis: vascular intrinsic and extrinsic factors and potential role of IL-6. *Nat Rev Cardiol* [Internet]. 2021 [citado 21 de diciembre de 2024];18(1):58-68. Disponible en: <https://www.nature.com/articles/s41569-020-0431-7>
163. Furlanetto M, Chagas E, Payao S. Atherosclerotic extension of carotid arteries: an insertion in clinical practice. *Int J Vasc Med* [Internet]. 2020 [citado 27 de junio de 2021];2020: 3120327. DOI: <https://doi.10.1155/2020/3120327>
164. Dawson L, Lum M, Nerleker N, Nicholls S, Layland J. Coronary atherosclerotic plaque regression: JACC state of the art review. *JACC* [Internet]. 2022 [citado 20 de agosto de 2023]; 79(1): 66- 82. Disponible en: <https://www.jacc.org/doi/epdf/10.1016/j.jacc.2021.10.035>
165. Omari M, Alkhalil M. Atherosclerosis Residual Lipid Risk-Overview of Existing and Future Pharmacotherapies. *J Cardiovasc Dev Dis* [Internet]. 2024 [citado 20 de diciembre de 2023];11(4):126. Disponible: <https://pubmed.ncbi.nlm.nih.gov/38667744/>
166. Vega Abascal JB, Piriz Assa AR, Nápoles Riaño DA, Baster Moro JC. ¿Podríamos ir más allá de la estrategia HEARTS para la predicción del riesgo cardiovascular en Cuba?

- Acta Médica [Internet]. 2024 [citado 4 de enero de 2025];25. Disponible en: <https://revactamedica.sld.cu/index.php/act/article/view/502>
167. Calvo-Betancourt LS, Ardila-Ardila DJ, García-Peña Á, Muñoz-Velandia OM, Ruiz ÁJ, Calvo-Betancourt LS, et al. Colesterol total en dislipidemias. ¿Útil medirlo? Acta Medica Colombiana [Internet]. 2019 [citado 30 de julio de 2020];44(4):11-3. Disponible en: http://www.scielo.org.co/scielo.php?script=sci_abstract&pid=S0120-24482019000400011&lng=en&nrm=iso&tlng=es
168. Aroundas A, Narayan S, Arnett D, Spector- Bagdady K, Bennett D, Cell L, et al. Use of artificial intelligence in improving outcomes in heart diseases: A scientific statement from the American Heart Association. Circulation [Internet]. 2024 [citado 10 de julio de 2024];149: e. DOI: <https://doi.org/10.1161/CIR0000000000001201>
169. Rubino F, Cummings D, Eckelf H, Coheng R, Wildingh J, Brown W, et al. Definition and diagnostic criteria of clinical obesity. The Lancet Diabetes & Endocrinology [Internet]. 2025 [citado 14 de enero de 2025]. Disponible en: [https://www.thelancet.com/journals/landia/article/PIIS2213-8587\(24\)00316-4/fulltext](https://www.thelancet.com/journals/landia/article/PIIS2213-8587(24)00316-4/fulltext)
170. Rubino F. Redefining obesity: advancing care for better lives. The Lancet Diabetes & Endocrinology [Internet]. 2025 [citado 14 de enero de 2025];0(0). Disponible en: [https://www.thelancet.com/journals/landia/article/PIIS2213-8587\(25\)00004-X/fulltext](https://www.thelancet.com/journals/landia/article/PIIS2213-8587(25)00004-X/fulltext)
171. Wah-Kheong C, Jacob G. Metabolic fatty liver syndromes: where do we stand in 2024. J Gastroenterology Hepatology [Internet]. 2024 [citado 12 de julio de 2024];39(4):613-614. DOI: <https://doi.org/10.1111/jgh.16507>
172. Deng L, Liu T, Liu CA, Zhang Q, Song MM, Lin SQ, et al. The association of metabolic syndrome score trajectory patterns with risk of all cancer types. Cancer [Internet]. 2024

[citado 12 de julio de 2024];130(12): 2150-9.DOI:
<http://doi.wiley.com/10.1002/cncr.35235>

173. Romeo CM, Lazcoz G. Inteligencia artificial aplicada a la salud: ¿qué marco jurídico? [Internet]. 2020 [citado 27 de junio de 2021]. Disponible en: <https://www.fundacionmercksalud.com/wp-content/uploads/2020/03/1.3.-IA-APLICADA-A-LA-SALUD.-Carlos-M.-Romeo-Guillermo-Lazcoz.pdf>
174. Van Smeden M, Heinze G, Calster B, Asselbergs F, Vardas P, Bruining N, et al. Critical appraisal of artificial intelligence-based prediction models for cardiovascular disease. *European Heart Journal* [Internet]. 2022 [citado 27 de junio de 2023];43. DOI: <http://dx.doi.org/10.1093/eurheartj/ehac238>
175. Chiarito M, Luceri L, Oliva A, Stefanini G, Condorelli G. Artificial Intelligence and Cardiovascular Risk Prediction: All That Glitters is not Gold. *Eur Cardiol* [Internet]. 2022 [citado 2 de enero de 2025];17: e29. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9947926/>
176. Teshale AB, Htun HL, Vered M, Owen AJ, Ryan J, Tonkin A, et al. Artificial intelligence improves risk prediction in cardiovascular disease. *GeroScience* [Internet]. 2024 [citado 2 de enero de 2025]. DOI: <https://doi.org/10.1007/s11357-024-01438-z>

ANEXOS

Anexo 1

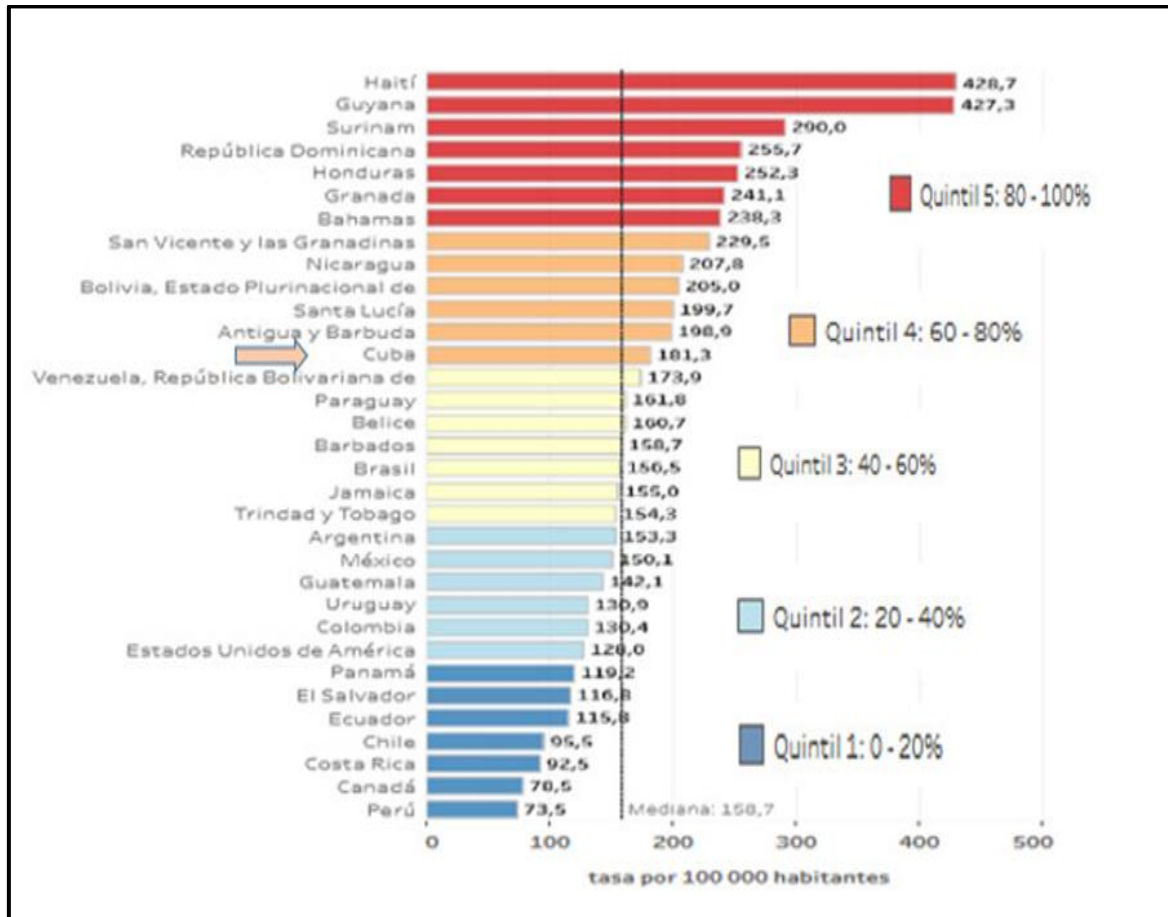


Gráfico 1. Mortalidad y carga de las enfermedades cardiovasculares en la Región de Las Américas, 2000-2019.

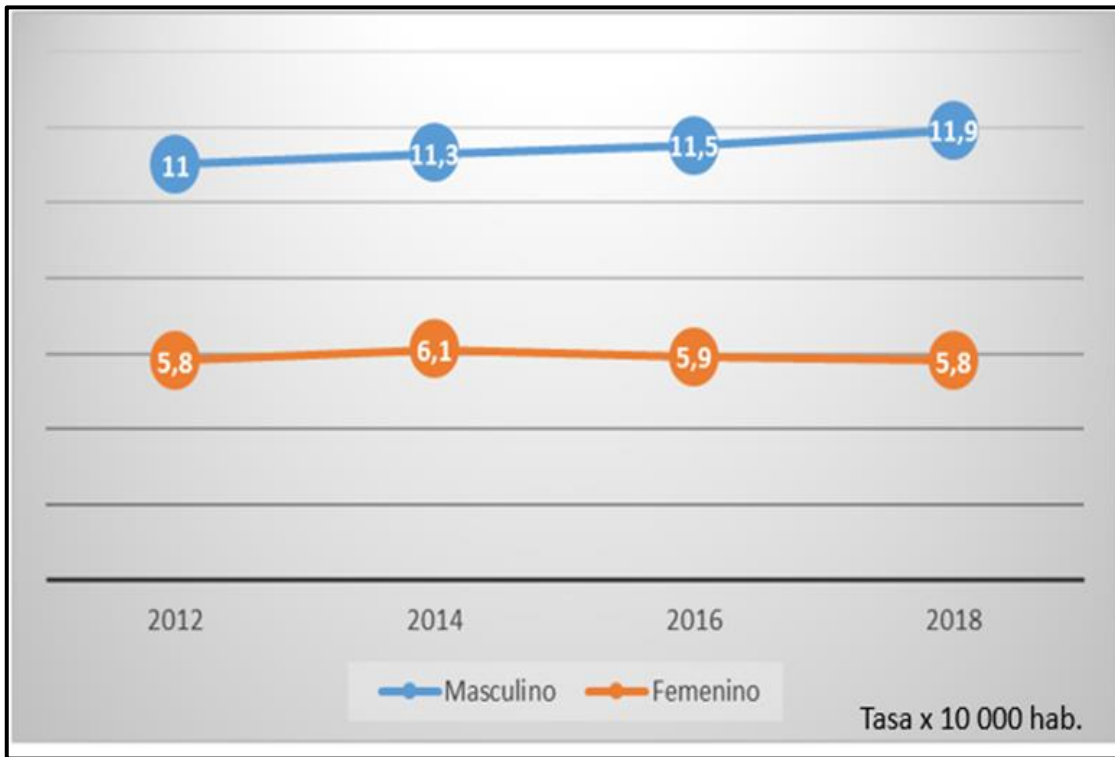


Gráfico 2. Mortalidad por infarto agudo del miocardio, en pacientes de 40 a 79 años en 2012, 2014, 2016 y 2018, Cuba.

Tabla 1. Factores de riesgo emergentes de enfermedades cardiovasculares.

Biomarcadores lipídicos
<ul style="list-style-type: none"> • Lipoproteína (a) • Apolipoproteína A1 y B • Lipoproteína asociada a fosfolipasa A2
Biomarcadores inflamatorios
<ul style="list-style-type: none"> • Proteína C reactiva • Interleucinas 1, 6 y 18 • Factor de necrosis tumoral α
Biomarcadores de hemostasia y trombosis
<ul style="list-style-type: none"> • Fibrinógeno • Factores de coagulación II, V y VIII • Antígeno del factor von Willebrand • Activador de plasminógeno tisular • Inhibidor del activador del plasminógeno 1 • Dímero D
Biomarcadores cardíacos
<ul style="list-style-type: none"> • Troponina de alta sensibilidad • Péptido natriurético tipo B
Biomarcadores renales
<ul style="list-style-type: none"> • Creatinina • Microalbuminuria • Cistatina C • Calcio • Fosfato/factor de crecimiento fibroblástico 23 • Ácido úrico
Otros factores
<ul style="list-style-type: none"> • Factores psicosociales • Técnicas de imagen

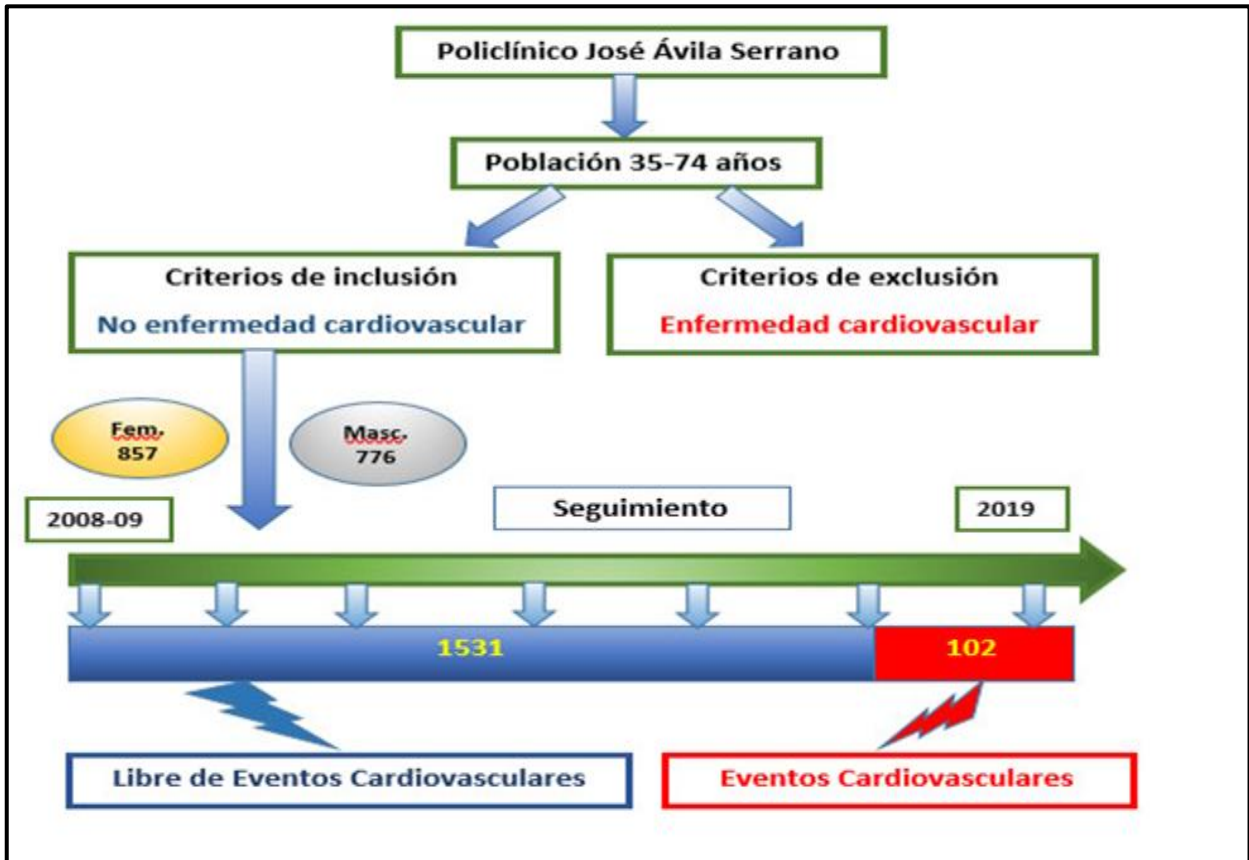


Figura 1. Ensamble y seguimiento de la cohorte prospectiva PredRCG_aps.

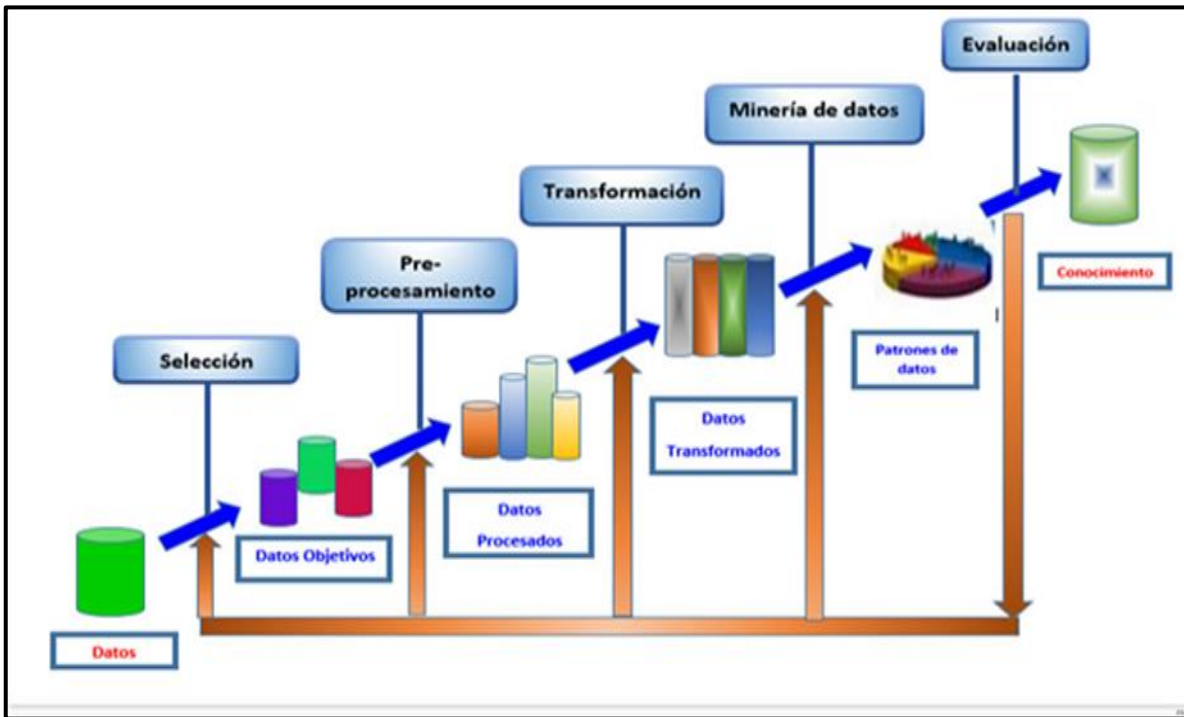


Figura 2. Etapas del proceso KDD.



Figura 3. Interfaz gráfica de Weka.

```
@relation riesgocardv1-weka.filters.supervised.instance.SMOTE-C0-K5-
P650.0-S1@attribute edad numeric@attribute sexo {No,Si}@attribute hfum
{No,Si}@attribute apphta {Si,No}@attribute appdm {No,Si}@attribute tas
numeric@attribute tad numeric@attribute imc numeric@attribute cirabd
numeric@attribute coltot numeric@attribute triglic numeric@attribute
colhdl numeric@attribute colldl numeric@attribute aciduric numeric
@attribute glicay numeric@attribute pcr-as numeric@attribute Evento
{No,Si}@data
63,No,No,Si,No,136,90,30.9,87,8.3,3.4,1.2,5.5,220,5.1,0.7,No
59,No,No,Si,No,160,95,30.6,124,4.3,2.2,1.1,2.2,310,9.5,1.5,No
55,No,No,Si,Si,140,100,30.1,98,5.5,2.8,1.2,3,220,9.5,1.7,No
44,No,No,Si,No,130,90,25.7,76,5.7,2.6,1.1,3.4,285,3.9,0.6,No
71,No,No,Si,No,140,90,25.3,85,5.3,0.5,1.3,3.8,400,3.9,0.8,No
39,No,No,Si,No,130,95,27.9,84,3.3,2.6,1.2,0.9,400,4.1,1,No
46,No,No,No,No,120,85,25.2,77,5.1,2.3,1,3,220,3.8,0.6,No
56,No,Si,No,No,130,90,29.4,78,5.5,1.5,1.2,3.6,296,8.2,0.7,No
48,No,No,Si,Si,140,85,30.2,86,4,1.7,1.1,2.1,285,7,2.6,No
64,No,No,Si,Si,150,80,22,75,5.7,1.7,1.1,3.8,250,6.4,1.6,No
55,No,No,Si,Si,130,80,29.3,100,4.7,1.6,1,2.9,250,7.4,0.9,No
68,No,No,No,Si,120,80,28.5,100,5.9,1.9,1.1,4,357,5.3,1,No
```

Figura 4. Fracción del archivo riesgocardv1.arff.

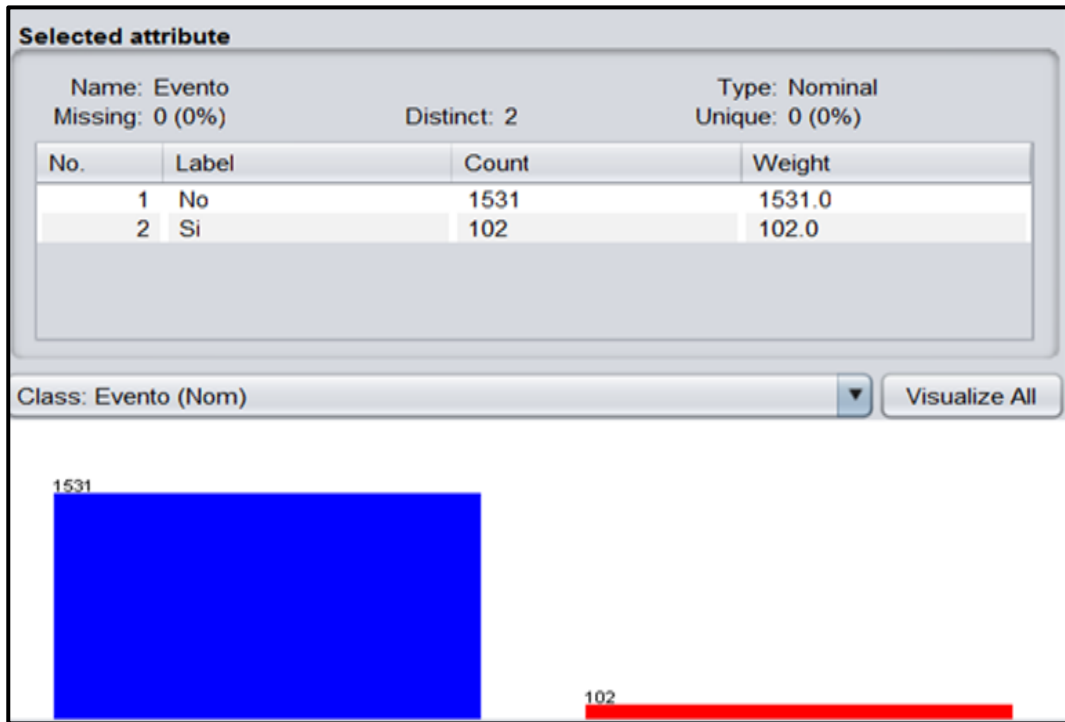


Figura 5. Clases desbalanceadas.

		Predicción		Total
		+	-	
Realidad	+	True Positive (TP)	False Negative (FN)	Positivos (+)
	-	False Positive (FP)	True Negative (TN)	Negativos (-)

Notas: **True Positive (TP):** verdaderos positivos; **True Negative (TN):** verdaderos negativos; **False negative:** falsos negativos; **False positive (FP):** falsos positivos

Figura 6. Matriz de confusión.

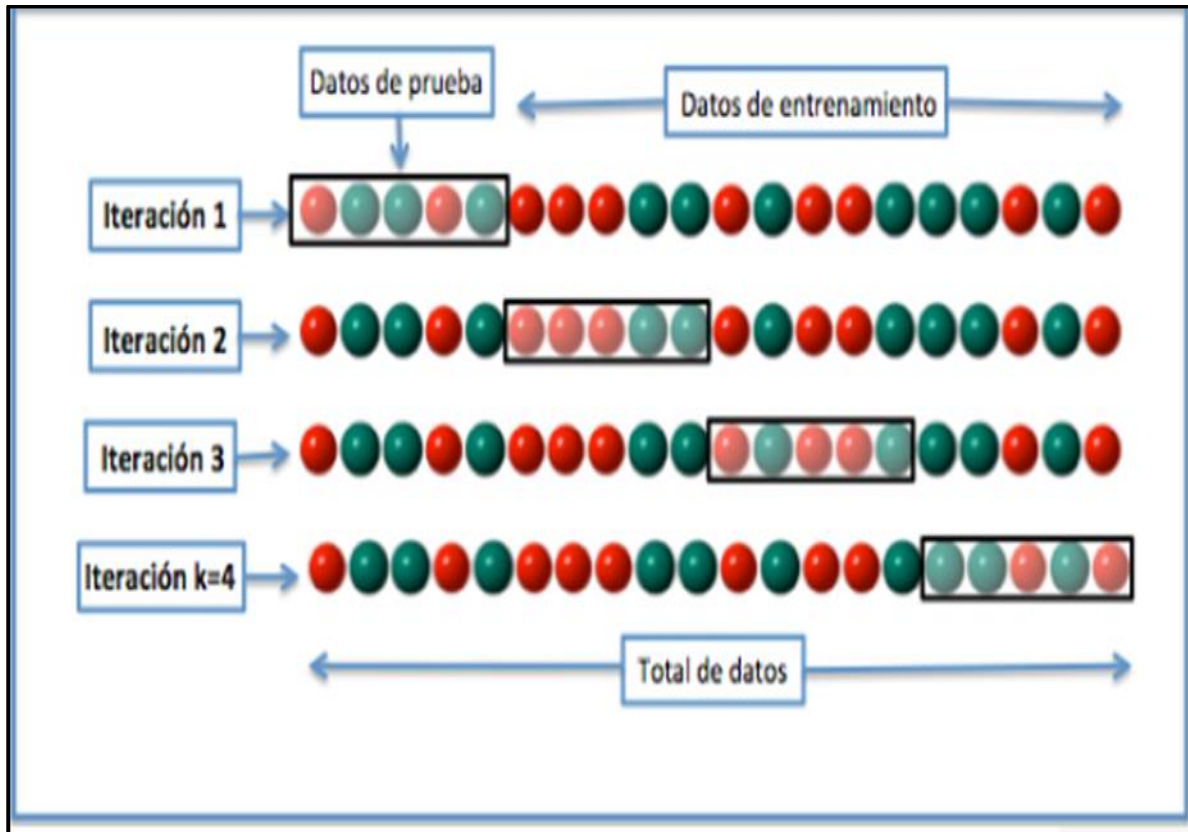


Figura 7. Técnica de validación cruzada.

ANEXO 2

Cuestionario para la recolección de la información a la inclusión y durante el seguimiento.

Policlínico Docente

José Ávila Serrano

Predicción del riesgo de enfermedad cardiovascular en la atención primaria de salud

Fecha D/M/A Inclusión_____ Seguimiento_____ Registro No_____

Nombre Y Apellidos: _____ edad _____ sexo_____

CI_____ Dirección_____ CMF No_____

APP

APF

Medicación: medicamentos y dosis día

Hábitos tóxicos

Anamnesis:

Examen Físico

Peso: talla: IMC circunferencia abdominal

tensión arterial sistólica: tensión arterial diastólica:

Frecuencia cardíaca:

Otros datos de interés:

Exámenes de laboratorio

colesterol total: colesterol HDL: colesterol LDL: triglicéridos:

glucemias ayunas: ácido úrico: proteína c reactiva as:

Otros exámenes.

Evento cardiovascular código CIE 10 Fecha:

Fuente:

Observaciones:

Salida seguimiento por otras causas:

ANEXO 3

Operacionalización de las variables del estudio de cohorte prospectivo

Variable dependiente (resultado)

Se consideró un evento cardiovascular mortal o no en el periodo de seguimiento, como: angina de pecho, infarto del miocardio, enfermedad cerebrovascular isquémica o hemorrágica, crisis de isquemia cerebral transitoria y enfermedad arterial periférica, se empleó la Clasificación Internacional de Enfermedades (CIE- 10), definiéndose las causas de muertes como enfermedades cardiacas (CIE-10: 120, 121-122, 124), enfermedades cerebrovasculares (CIE-10: 160-162,163,164) y otras enfermedades cardiovasculares(CIE-10: 173).

Cualitativa nominal dicotómica: evento cardiovascular (sí/no)

Variables independientes

edad: en años cumplidos.

Cuantitativa continua expresada en años.

sexo: en masculino o femenino según fenotipo

Cualitativa nominal dicotómica.

hfumar: se refiere al consumo de tabaco.

Cualitativa nominal dicotómica (sí/no).

apphta: referida a los antecedentes de hipertensión arterial.

Cualitativa nominal dicotómica (sí/no).

appdm: Diabetes mellitus previamente diagnosticada o nivel de glucosa en ayunas > 7 mmol /L o > 11.1 mmol /L a las 2 horas.

Cualitativa nominal dicotómica (sí/no).

IMC: Índice de Masa Corporal, fue calculado dividiendo el peso en kg sobre la talla en metros al cuadrado.

Cuantitativa continúa expresado en Kg/m².

circabd: Se realizó la medición de la circunferencia abdominal en la intercepción de la línea axilar media y el borde superior de la cresta iliaca, con una cinta métrica perpendicular al eje longitudinal del cuerpo con la persona de pie.

Cuantitativa continúa expresada en cm.

tas (tensión arterial sistólica): Para la determinación de la presión arterial se siguieron las recomendaciones de la guía cubana de prevención, diagnóstico y tratamiento de la hipertensión arterial. Se determinó la presión arterial en el brazo derecho con el paciente sentado, usando un esfigmomanómetro anerode en 2 ocasiones, el promedio de ambas mediciones se consideró la tensión arterial.

Cuantitativa continúa expresada en mmHg.

tad (Tensión arterial diastólica): Se siguió el mismo procedimiento usado para la tensión arterial sistólica.

Cuantitativa continúa expresada en mmHg.

coltot (Colesterol total): Se determinó el nivel de colesterol en sangre usando reactivo Colestest de producción nacional (Laboratorios Finlay, Ciudad de La Habana).

Cuantitativa continúa expresada en mmol/L.

cHDL (colesterol unido a lipoproteínas de alta densidad): Se determinó por método homogéneo de C-HDL Inmuno FS, de producción nacional (Laboratorios Finlay, Ciudad de La Habana).

Cuantitativa continúa expresada en mmol/L.

cLDL (colesterol unido a proteína de baja densidad): Se determinó con el reactivo de Friedewald o se calculó, si procedía, según la fórmula de Friedewald, basado en los niveles de colesterol total, colesterol HDL y triglicéridos:

$$\text{LDL-C} = \text{Colesterol total} - [\text{TAG}/2,1] - \text{HDL-C}$$

Cuantitativa continua expresada en mmol/L.

triglic: (**triglicéridos**) se determinó usando reactivo Triglitest de producción nacional (Laboratorios Finlay, Ciudad de La Habana).

Cuantitativa continua expresada en mmol/L.

aciduric (ácido úrico): Se determinó mediante el reactivo uric acid mono SI, de producción nacional (Laboratorios Finlay, Ciudad de La Habana).

Cuantitativa continua expresada en mg/dL.

glicay (glucemia en ayunas): Se determinó con reactivo Glucotest, por método glucosa oxidasa (GOD), utilizando reactivo de producción nacional (Laboratorios Finlay, Ciudad de La Habana).

Cuantitativa continua expresado en mmol/L.

pcr-as o pcr-hs (Proteína C reactiva de alta sensibilidad): Fue determinada en 2 ocasiones con un intervalo de 15 días entre las mediciones, utilizando el valor promedio de ambas, se usó el reactivo anti-CRP, este reactivo contiene anticuerpos específicos que se unen a la proteína c reactiva, midiéndose turbidimétricamente, y se detectaron concentraciones de hasta 0,01 mg/dL.

Cuantitativa continua expresado en mg/dL.

Tabla 2. Descripción de los atributos predictores en Weka.

No	Nombre del atributo	Descripción	Formato o fuente
1	edad	Identifica la edad del paciente	Numérico
2	sexo	Identifica sexo de paciente	femenino, masculino
3	apphta	Identifica antecedente de hipertensión arterial	Sí_No
4	appdm	Identifica antecedente de diabetes mellitus	Sí_No
5	hfumar	Identifica el consumo de tabaco	Sí_No
6	tas	Identifica cifras de tensión arterial sistólica	Numérico
7	tad	Identifica cifras de tensión arterial diastólica	Numérico
8	imc	Identifica el Índice de Masa Corporal	normopeso, sobrepeso, obeso
9	circabd	Identifica la medida de circunferencia abdominal (Sí>92 cm hombres,>88 cm mujer, resto No)	Sí_No
10	coltot	Identifica nivel de colesterol total	Numérico
11	cHDL	Identifica nivel de colesterol HDL	Numérico
12	cLDL	Identifica nivel de colesterol LDL	Numérico
13	triglic	Identifica nivel de triglicéridos	Numérico
14	glicay	Identifica nivel de glucemia en ayunas	Numérico
15	aciduric	Identifica nivel de ácido úrico	Numérico
16	pcr-as	Identifica nivel de proteína c reactiva de alta sensibilidad	Numérico

Anexo 4. Métricas para clases desbalanceadas.

Dos medidas importantes son la **tasa de verdaderos positivos (TPrate)** y la de **falsos positivos (FPrate)**, calculadas como:

Tasa de verdaderos positivos (TPrate): es el porcentaje de instancias positivas correctamente clasificadas. **(Ecuación 1)**

$$\text{Tasa de verdaderos positivos} = \frac{\text{TP}}{\text{TP+FN}} \quad (1)$$

Tasa de falsos positivos (FPrate): es el porcentaje de instancias negativas incorrectamente clasificadas como positivas. **(Ecuación 2)**

$$\text{Tasa de falsos positivos} = \frac{\text{FP}}{\text{FP+TN}} \quad (2)$$

Otras métricas utilizadas en clases desbalanceadas son:

F-measure (medida F): se utiliza para integrar la **Precisión y Recall** dentro de una métrica, lo que representa una media armónica ponderada entre estos dos parámetros.

(Ecuación 3)

$$F - Measure = \frac{(1+\beta)^2 \cdot Recall \cdot Precision}{(\beta^2 \cdot Recall + Precision)}$$

Donde β es un coeficiente para ajustar la importancia relativa de la Precisión contra el Recall (usualmente $\beta = 1$)

Siendo la **precisión** el porcentaje de predicciones positivas que son correctas, y el **recall (cubrimiento)**, el porcentaje de verdaderos positivos que son correctamente detectados, o razón de verdaderos positivos, el cubrimiento es conocido también como exactitud de la clase positiva o **sensibilidad (sensitivity)**.

Las curvas PRC (Precision-Recall Curve) pueden ofrecer mayor información sobre la valoración del desempeño en el caso de conjuntos de datos desbalanceados; por esto muchos trabajos actuales usan este tipo de curvas para evaluaciones de desempeño y comparaciones, estas curvas se definen graficando la tasa de Precisión contra la tasa de Recall.

Las **curvas PR** tienen una estrecha relación con las **curvas ROC**, una curva domina en el espacio ROC sí, solo sí, domina en el espacio PR.

Curvas ROC (Receiver Operating Characteristics): se forma graficando **TPrate** contra **FPrate**, por lo que un punto en este espacio corresponde al desempeño de un algoritmo de clasificación en una distribución dada, esta curva es muy útil, pues ofrece

una representación visual del **compromiso (trade-off)** entre los **beneficios** (reflejados por los TP) y los **costos** (reflejados por los FP) de la clasificación.

Ha sido adoptada por la comunidad de investigadores en minería de datos y es un indicador de la calidad del modelo, se pueden interpretar las curvas ROC utilizando los intervalos de valores del área bajo la curva, (*Area Under Curve*, AUC) de la siguiente guía:

[0.5 – 0.6): Test malo

[0.6 – 0.75): Test regular

[0.75 – 0.9): Test bueno

[0.9 – 0.97): Test muy bueno

[0.97 – 1]: Test excelente

Coefficiente de Correlación de Matthews, (MCC, en inglés): mide la efectividad global de un clasificador, es recomendado para clases desbalanceadas, se usa como una medida de la calidad de las clasificaciones de dos clases, devuelve un valor entre -1 (ninguna relación entre predicción y observación) y 1 (predicción perfecta). **(Ecuación 4)**

$$MCC = \frac{TP.TN - FP.FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

(4)

Anexo 5. Consentimiento informado.

Policlínico Docente

“José Ávila Serrano”

La investigación sobre la predicción del riesgo de enfermedades cardiovasculares basado en técnicas de inteligencia artificial en la atención primaria de salud reviste gran importancia, por la elevada morbilidad y mortalidad de estas enfermedades en la población. El propósito de esta investigación es el diseño de un modelo predictivo para predecir eficazmente al paciente asintomático con riesgo de desarrollar una enfermedad cardiovascular y por consiguiente mejorar la predicción, prevención y tratamiento de estas enfermedades en la atención primaria de salud, como parte de una investigación doctoral.

El presente documento de consentimiento informado, entregado por el investigador del trabajo titulado “Modelo predictivo de enfermedad cardiovascular basado en inteligencia artificial para la atención primaria de salud”, previa consulta oral, es parte del protocolo de investigación que ha sido aprobado por el Comité de Ética correspondiente.

Yo _____ CI: _____

Ejerciendo mi libre poder de elección y mi voluntad expresa, por este medio, doy mi consentimiento para participar en el estudio. También se me ha explicado que la información que yo brinde es de carácter confidencial, no serán divulgados ni publicados, ni mi identidad, ni los detalles personales. Estoy consciente de mi derecho a no responder cualquier pregunta que considere indiscreta. Se me ha explicado que la

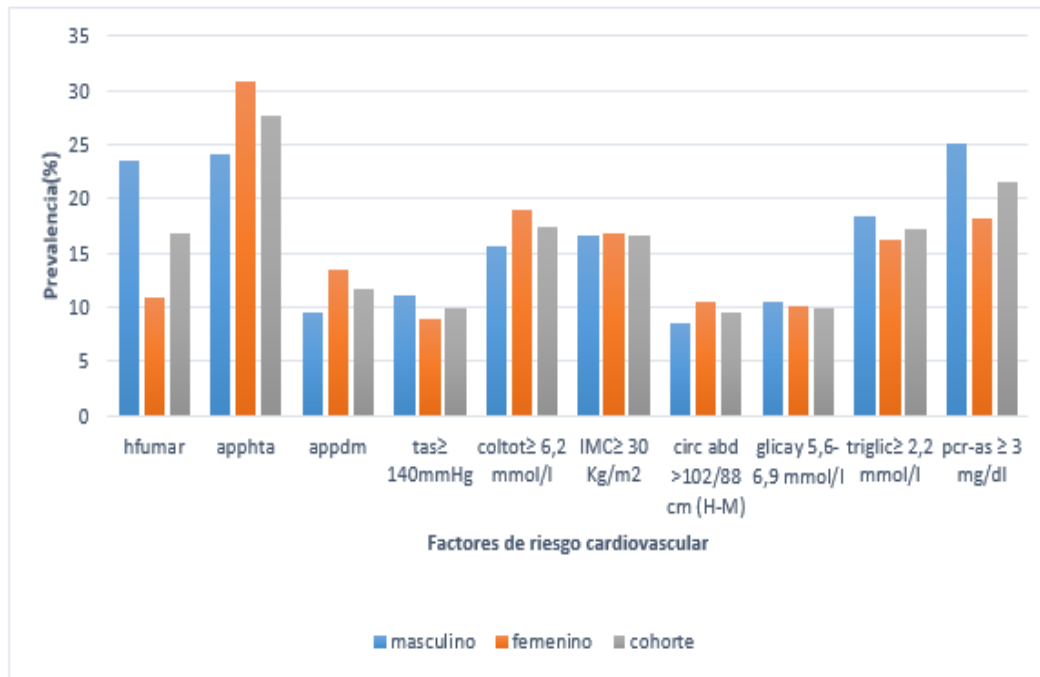
investigación se efectuará en el contexto de la atención primaria de salud y no será privado del protocolo de atención establecidos, sin que afecte las relaciones con el equipo médico aún si me niego a participar en el estudio.

Para que así conste firmo el presente consentimiento a los _____ días del mes _____ del 20____.

Firma. _____

Investigador. _____

Anexo 6. Resultados y discusión.



Notas: hfumar: hábito de fumar, apphta: antecedentes personales de hipertensión arterial, appdm: antecedentes personales de diabetes mellitus, tas: tensión arterial sistólica, coltot: colesterol total, IMC: Índice de Masa Corporal, circabd ≥ 102/88 cm hombre y mujer respectivamente, glicay: glicemia en persona no diabéticas, triglic: triglicéridos, pcr-as: proteína c reactiva de alta sensibilidad.

Gráfico 3 Prevalencia de factores de riesgo cardiovascular seleccionados según sexos y en la cohorte.

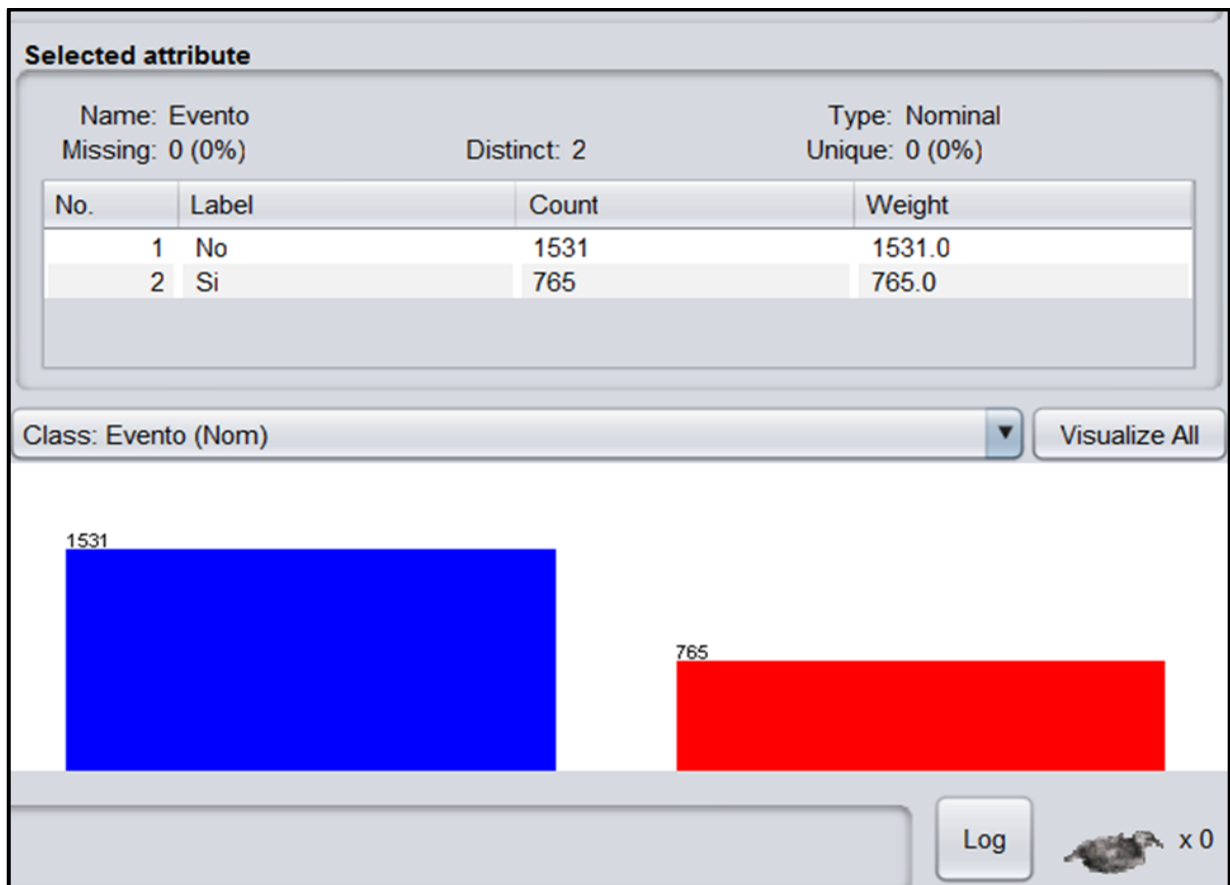


Figura 8. Datos balanceados con el algoritmo SMOTE.

Tabla 3. Codificación de los atributos predictores en Weka.

No	Atributos	Descripción	Codificación Weka
1	edad	Identifica los años cumplidos	Numérico
2	sexo	Identifica el sexo del paciente	femenino/masculino
3	hfum	Identifica el consumo de tabaco	Sí_No
4	apphta	Antecedentes de hipertensión	Sí_No
5	appdm	Antecedentes de diabetes mellitus	Sí_No
6	tas	Cifras de tensión arterial sistólica	
		<140 mm Hg	Optima
		140-159 mm Hg	Estadio I
		≥ 160 mm Hg	Estadio II
7	tad	Cifras de tensión arterial diastólica	
		<90 mm Hg	Optima
		90-99 mm Hg	Estadio I
		≥ 100 mm Hg	Estadio II
8	imc	Índice de masa corporal	
		18,5-24,9 Kg/m ²	Normopeso
		25- 29,9 Kg/m ²	Sobrepeso
		≥ 30 Kg/m ²	Obeso
9	circabd	Circunferencia abdominal	
		Masculino ≥ 102 cm /Femenino ≥ 88 cm	Sí_No
10	coltot	Cifras de colesterol total	
		< 6 mmol/L	Deseable
		6-7,2 mmol/L	Riesgo moderado
		> 7,2 mmol/L	Alto riesgo

11	cHDL	Cifras de colesterol lipoproteína de alta densidad	
		Fem < 1,3 mmol/L / Masc.< 1 mmol/L	Sí_No
12	cLDL	Cifras de colesterol lipoproteínas de baja densidad	
		< 3,37 mmol/L	Deseable
		3,38- 4,15 mmol/L	Riesgo moderado
		≥ 4,16 mmol/L	Riesgo alto
13	triglic	Cifras de triglicéridos	
		< 1,7 mmol/L	Deseable
		1,7-2,2 mmol/L	Riesgo moderado
		≥ 2,3 mmol/L	Riesgo alto
14	glicay	Cifras de glucemia en ayunas	
		< 5,6 mmol/L	Deseable
		5,6-6,9 mmol/L	Riesgo moderado
		≥ 7 mmol/L	Riesgo alto
15	aciduric	Cifras de ácido úrico	
		Masc. > 428 mmol/L / Fem. > 357mmol/L	Sí_No
16	pcr-as	Cifras de proteína c reactiva de alta sensibilidad	
		< 1mg/dL	Deseable
		1-3 mg/dL	Riesgo moderado
		> 3 mg/dL	Riesgo alto

Tablas 4. Resultados de los métodos de selección de atributos.

Atributos	Métodos de selección			
	CfsSubsetEval /BestFirst	WrapperSubset Eval/ J48/BestFirst	CorrelationAttributeEval/ Ranker	
			Merito promedio	Rango promedio
edad	10%	50%	0,228+-0,008	7+- 0,45
sexo	0%	50%	0,058+-0,006	15,4+-0,49
hfum	100%	90%	0,257+-0,009	6,1+-0,3
apphta	100%	40%	0,329+-0,005	1,8+-0,4
appdm	100%	20%	0,295+-0,012	4,5+-0,81
tas	100%	10%	0,305+-0,007	3,7+-0,78
tad	0%	0%	0,178+-0,006	12,9+-0,03
IMC	0%	30%	0,055+-0,004	15,6+-0,49
circabd	100%	90%	0,339+-0,009	1,2+-0,4
coltot	40%	40%	0,195+-0,004	10,1+-0,54
triglic	100%	30%	0,218+-0,005	7,9+-0,3
cHDL	100%	50%	0,202+-0,006	9,1+-0,3
cLDL	60%	20%	0,188+-0,007	11,1+-0,7
aciduric	0%	30%	0,122+-0,01	14
glicay	100%	70%	0,305+-0,008	3,8+-0,6
pcr-as	100%	60%	0,186+-0,003	11.8+-0,6

Tabla 5. Reglas obtenidas con el algoritmo JRip.

No	Reglas
r1	Sí ... (cLDL \geq 4,16 mmol/L) and (apphta = Sí) and (triglic \geq 2,3 mmol/L) and (pcr-as > 3 mg/dL) => Evento= Sí (491.0/4.0)
r2	Sí ... (apphta = Sí) and (cLDL \geq 4,16 mmol/L) and (edad \geq 58) and (cHDL <1,3/1 mmol/L) and (pcr-as = 1-3 mg/dL) => Evento= Sí (96.0/5.0)
r3	Sí ... (pcr-as > 3 mg/dL) and (apphta = Sí) and (tas = 140-159 mmHg) and (triglic \geq 2,3 mmol/L) and (edad \leq 59) => Evento= Sí (41.0/0.0)
r4	Sí (pcr-as > 3 mg/dL) and (glicay = 5,6-7.2 mmol/L) and (cLDL \geq 4,16 mmol/L) and (edad \geq 53) => Evento= Sí (33.0/4.0)
r5	Sí (pcr-as > 3 mg/dL) and (edad \geq 71) and (coltot \geq 7,2 mmol/L) => Evento= Sí (47.0/2.0)
r6	Sí ... (pcr-as > 3 mg/dL) and (cHDL) <1,3 /1 mmol/L) and (sexo = masculino) and (apphta = Sí) and (imc = Sobrepeso) => Evento= Sí (25.0/4.0)
r7	Sí ... (edad \geq 68) and (triglic \geq 2,3 mmol/L) and (tad = 90-99 mmHg) and (cHDL<1,3 /1 mmol/L) and (imc = Sobrepeso) => Evento= Sí (13.0/0.0)
r8	Sí ... (pcr- as > 3 mg/dL) and (edad \geq 56) and (triglic = 1,7-2,2 mmol/L) => Evento= Sí (9.0/2.0)

Classifier Model

J48 pruned tree

apphta = Si

| colhdl = Si

| | coldl = Alto riesgo

| | | pcrhs = Deseable: No (16.0/1.0)

| | | pcrhs = Riesgo Mod

| | | | tas = Optimo: No (8.0/1.0)

| | | | tas = Estadio 2: Si (13.0/2.0)

| | | | tas = Estadio 1

| | | | | hfum = No

| | | | | | cirabd = No: No (4.0)

| | | | | | cirabd = Si: Si (15.0)

| | | | | hfum = Si: Si (65.0/1.0)

| | | | pcrhs = Alto riesgo: Si (579.0/17.0)

| | | coldl = Deseable

| | | | hfum = No

| | | | | cirabd = No: No (73.0)

| | | | | cirabd = Si

| | | | | | pcrhs = Deseable: No (3.0)

| | | | | | pcrhs = Riesgo Mod: No (5.0)

| | | | | | pcrhs = Alto riesgo: Si (14.0/2.0)

| | | | hfum = Si

| | | | | pcrhs = Deseable: No (2.0)

| | | | | pcrhs = Riesgo Mod: No (9.0/1.0)

| | | | | pcrhs = Alto riesgo: Si (37.0/5.0)

Figura 9. Árbol de decisión (fracción) generado por el meta-algoritmo de clasificación Attribute Selected Classifier junto con el algoritmo J 48.

Tabla 6. Reglas obtenidas con el algoritmo J48.

No	Reglas
r1	Sí ... (apphta = Sí) and (cHDL) <1,3 /1 mmol/L) and (cLDL >= 4,16 mmol/L) and (pcr-as > 3 mg/dL) => Evento= Sí (579,0/17,0)
r2	Sí ... (apphta = Sí) and (cHDL) <1,3 /1 mmol/L) and (cLDL >= 4,16 mmol/L) and (pcr-as 1-3 mg/dL) and (tas >= 160 mmHg) => Evento= Sí (13,0/2,0)
r3	Sí ... (apphta = Sí) and (cHDL) <1,3 /1 mmol/L) and (cLDL >= 4,16 mmol/L) and (pcr-as 1-3 mg/dL) and (tas = 140-159 mmHg) and (hfum= Sí) => Evento= Sí (65,0/1,0)
r4	Sí ... (apphta = Sí) and (cHDL) <1,3/1 mmol/L) and (cLDL >= 4,16 mmol/L) and (pcr-as 1-3 mg/dL) and (tas = 140-159 mmHg) and (hfum= Sí) and (circabd > 102/88 cm) => Evento= Sí (15,0/0,0)
r5	Sí ... (apphta = Sí) and (cHDL) <1,3/1 mmol/L) and (cLDL= 3,38- 4,15 mmol/L) and (tas = 140-159 mmHg) and (hfum= Sí) and (glicay=5,6-7,2 mmol/L) => Evento= Sí (16,0/0,0)
r6	Sí ... (apphta = Sí) and (cHDL) <1,3 /1 mmol/L) and (cLDL= 3,38- 4,15 mmol/L) and (tas = 140-159 mmHg) and (hfum= Sí) and (triglic >= 2,3 mmol/L) => Evento= Sí (9,0/1,0)
r7	Sí ... (apphta = Sí) and (cHDL) <1,3/1 mmo Sí l/l) and (cLDL <3,37 mmol/L) and (hfum= Sí) and (pcr-as > 3 mg/dL) => Evento= (37,0/5,0)
r8	Sí ... (apphta = Si) and (cHDL) <1,3/1 mmol/L) and (cLDL <3,37 mmol/L) and (hfum=No) and (circabd > 102/88 cm) and (pcr-as > 3 mg/dL) => Evento= Sí (14,0/2,0)

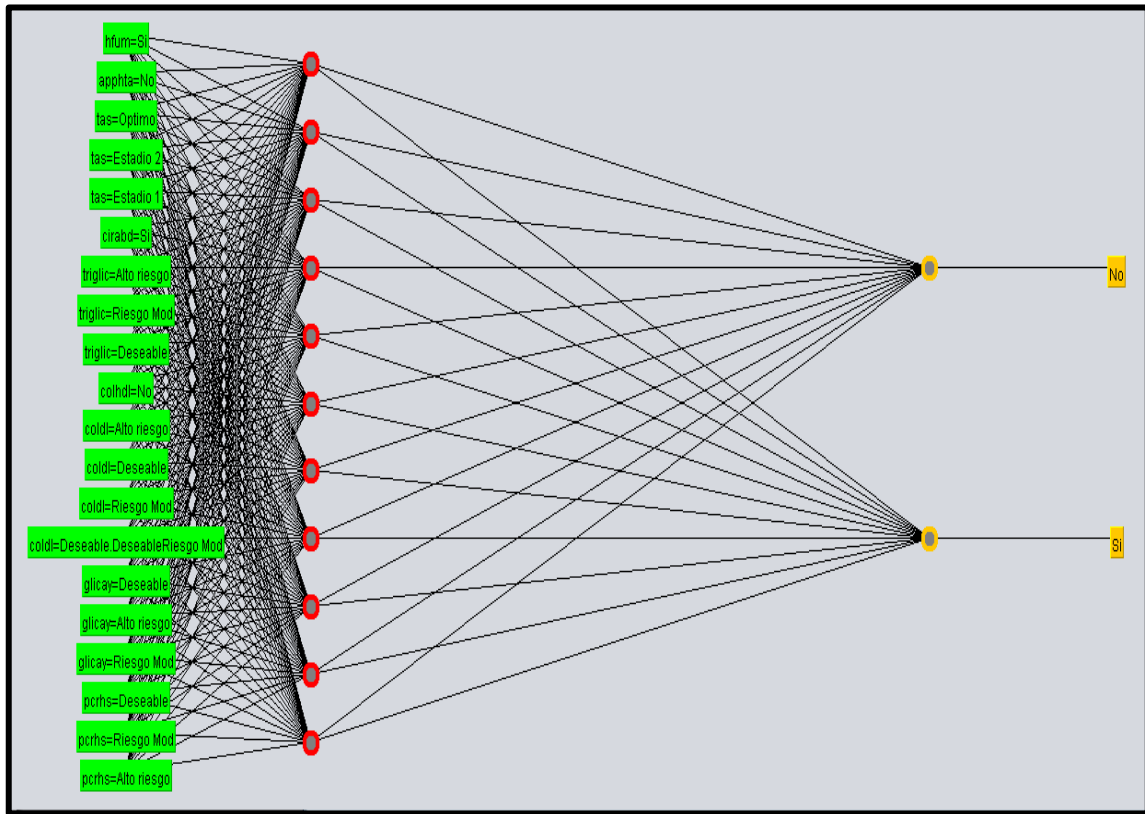


Figura 10. Arquitectura de la red neuronal.

Sigmoid Node 0	
Inputs	Weights
Threshold	-2.7415275293639616
Node 2	1.9158079197094384
Node 3	2.103401679789782
Node 4	5.5963472446532245
Node 5	3.554953886747977
Node 6	-3.5278194074294436
Node 7	2.562009225736332
Node 8	-4.187389145110488
Node 9	-4.990308663946565
Node 10	3.1330445556359736
Node 11	3.675603168437151
Node 12	-1.5058728520519398
Node 13	2.7426712548701695
Node 14	5.577445850218778
Node 15	3.3923664589423743
Node 16	3.987190971631429
Node 17	-1.8964426121805351
Node 18	5.862314265374427

Sigmoid Node 1	
Inputs	Weights
Threshold	2.7450135137258243
Node 2	-1.913470564890451
Node 3	-2.1082298061418396
Node 4	-5.592748497247083
Node 5	-3.551805511871667
Node 6	3.5182398699060697
Node 7	-2.577644137064577
Node 8	4.189791847716598
Node 9	4.989885003757372
Node 10	-3.1339009160017484
Node 11	-3.671028730446808
Node 12	1.5618882102312206
Node 13	-2.7442441897120515
Node 14	-5.580159117295669
Node 15	-3.402099074359115
Node 16	-3.989717169909045
Node 17	1.844161218020613
Node 18	-5.853639730406259

Figura 11. Nodos 0 y 1 de la red neuronal Multilayer Perceptron.

Tabla 7. Evaluación del desempeño de los algoritmos clasificadores en los datos de entrenamiento.

Algoritmo clasificador	TPRate	FPRate	Curva ROC	F Measure	Curva PRC	MCC
<i>JRip</i>	0,959	0,014	0,977	0,966	0,964	0,949
<i>AttributeSelected Classifier/ J48</i>	0,942	0,020	0,951	0,994	0,949	0,927
<i>AttributeSelectedClassifier/Multilayer Perceptron</i>	0,987	0,010	0,993	0,993	0,995	0,990

Notas: TPRate: Tasa de verdaderos positivos; FPRate: Tasa de falsos positivos; Curva ROC: Área bajo la curva Receiver Operating Characteristics; F Measure: Medida F (FScore); Curva PRC: Curva Precisión_Recall; MCC: Coeficiente de Correlación de Matthews.

Tabla 8. Validación cruzada de los algoritmos clasificadores.

Algoritmo clasificador	TPRate	FPRate	Curva ROC	F Measure	Curva PRC	MCC
<i>JRip</i>	0,948	0,036	0,965	0,939	0,942	0,907
<i>AttributeSelected Classifier/ J48</i>	0,935	0,025	0,957	0,941	0,909	0,913
<i>AttributeSelectedClassifier/Multilayer Perceptron</i>	0,953	0,025	0,989	0,952	0,997	0,928

Notas: TPRate: Tasa de verdaderos positivos; FPRate: Tasa de falsos positivos; Curva ROC: Área bajo la curva Receiver Operating Characteristics; F Measure: Medida F (FScore); Curva PRC: Curva Precisión_Recall; MCC: Coeficiente de Correlación de Matthews.

PROBAST

Wolff et al, 2019 Annals Intern Med

PROBAST

A Tool to assess the Risk of Bias and Applicability of Prediction Model Studies

Domain / Key questions	Vega-Abascal, 2023	
0. Study information		
0.1 Author	Vega Abascal J	
0.2 Publication year	2023	
0.3 Publication identifier (Title, PMID, DOI)	Predictive model of cardiovascular disease based in artificial intelligence in primary health care	
0.4 Publication Journal	Revista Cubana de Medicina General Integral	
0.5 Model name	ModPredIA EnfCardiovasc_aps 1.0	
DOMAIN 1: Participants		
A. Risk of Bias		
	Dev	Val
1.1 Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?	Yes	
1.2 Were all inclusions and exclusions of participants appropriate?	Yes	
Risk of bias introduced by selection of participants	Low RoB	
B. Applicability	Low concern	
DOMAIN 2: Predictors		
A. Risk of Bias		
2.1 Were predictors defined and assessed in a similar way for all participants?	Yes	
2.2 Were predictor assessments made without knowledge of outcome data?	Yes	
2.3 Are all predictors available at the time the model is intended to be used?	Yes	
Risk of bias introduced by predictors or their assessment	Low RoB	
B. Applicability	Low concern	
DOMAIN 3: Outcome		

A. Risk of Bias		
3.1 Was the outcome determined appropriately?	Yes	
3.2 Was a pre-specified or standard outcome definition used?	Yes	
3.3 Were predictors excluded from the outcome definition?	Yes	
3.4 Was the outcome defined and determined in a similar way for all participants?	Yes	
3.5 Was the outcome determined without knowledge of predictor information?	Yes	
3.6 Was the time interval between predictor assessment and outcome determination appropriate?	Yes	
Risk of bias introduced by the outcome or its determination	Low RoB	
B. Applicability	Low concern	
DOMAIN 4: Analysis		
Risk of Bias		
4.1 Were there a reasonable number of participants with the outcome?	Yes	
4.2 Were continuous and categorical predictors handled appropriately?	Yes	
4.3 Were all enrolled participants included in the analysis?	Yes	
4.4 Were participants with missing data handled appropriately?	Yes	
4.5 Was selection of predictors based on univariable analysis avoided?	NA	
4.6 Were complexities in the data (e.g. censoring, competing risks, sampling of controls) accounted for appropriately?	Yes	
4.7 Were relevant model performance measures evaluated appropriately?	Yes	
4.8 Were model overfitting and optimism in model performance accounted for?	Yes	
4.9 Do predictors and their assigned weights in the final model correspond to the results from multivariable analysis?	NA	
4.1 Were there a reasonable number of participants with the outcome?	Yes	
Risk of bias introduced by the analysis	Low RoB	
Overall assessment		
Reaching an overall judgement about risk of bias of the prediction model evaluation		
Low risk of bias	<p>If all domains were rated low risk of bias.</p> <p>If a prediction model was developed without any external validation, and it was rated as low risk of bias for all domains, consider downgrading to high risk of bias. Such a model can only be considered as low risk of bias, if the development was based on a very large data set and included some form of internal validation.</p>	

Reaching an overall judgement about applicability of the prediction model evaluation	
Low concerns regarding applicability	If low concerns regarding applicability for all domains, the prediction model evaluation is judged to have low concerns regarding applicability
Overall judgement about risk of bias and applicability of the prediction model evaluation	
Overall judgement of risk of bias	Low RoB
Overall judgement of applicability	Low concern

Nota NA: not applicable.

Figura 12. PROBAST: Herramienta para evaluar el riesgo de sesgo y la aplicabilidad de los modelos predictivos.

Section /Topic	Item	D/E ¹	Checklist item	Reported on page
TITLE				
Title	1	D;E	Identify the study as developing or evaluating the performance of a multivariable prediction model, the target population, and the outcome to be predicted	0
ABSTRACT				
Abstract	2	D;E	See TRIPOD+AI for Abstracts checklist	Synthesis
INTRODUCTION				
Background	3a	D;E	Explain the healthcare context (including whether diagnostic or prognostic) and rationale for developing or evaluating the prediction model, including references to existing models	1-10, 11-14, 29-31
	3b	D;E	Describe the target population and the intended purpose of the prediction model in the context of the care pathway, including its intended users (e.g., healthcare professionals, patients, public)	1-10
	3c	D;E	Describe any known health inequalities between sociodemographic groups	NA
Objectives	4	D;E	Specify the study objectives, including whether the study describes the development or validation of a prediction model (or both)	6
METHODS				
Data	5a	D;E	Describe the sources of data separately for the development and evaluation datasets (e.g., randomized trial, cohort, routine care or registry data), the rationale for using these data, and representativeness of the data	38
	5b	D;E	Specify the dates of the collected participant data, including start and end of participant accrual; and, if applicable, end of follow-up	40
Participants	6a	D;E	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including the number and location of centers	38
	6b	D;E	Describe the eligibility criteria for study participants	39
	6c	D;E	Give details of any treatments received, and how they were handled during model development or	42
Data preparation	7	D;E	Describe any data pre-processing and quality checking, including whether this was similar across relevant sociodemographic groups	44-45
Outcome	8a	D;E	Clearly define the outcome that is being predicted and the time horizon, including how and when assessed, the rationale for choosing this outcome, and whether the method of outcome assessment is consistent across sociodemographic groups	41-42
	8b	D;E	If outcome assessment requires subjective interpretation, describe the qualifications and demographic characteristics of the outcome assessors	NA
	8c	D;E	Report any actions to blind assessment of the outcome to be predicted	42
Predictors	9a	D;E	Describe the choice of initial predictors (e.g., literature, previous models, all available predictors) and any pre-selection of predictors before model building	42-43

	9b	D;E	Clearly define all predictors, including how and when they were measured (and any actions to blind assessment of predictors for the outcome and other predictors)	40-41
	9c	D;E	If predictor measurement requires subjective interpretation, describe the qualifications and demographic characteristics of the predictor assessors	NA
Sample size	10	D;E	Explain how the study size was arrived at (separately for development and evaluation), and justify that the study size was sufficient to answer the research question. Include details of any sample size calculation	38-39
Missing data	11	D;E	Describe how missing data were handled. Provide reasons for omitting any data	42
Analytical methods	12a	D	Describe how the data were used (e.g., for development and evaluation of model performance) in the analysis, including whether the data were partitioned, considering any sample size requirements	46
	12b	D	Depending on the type of model, describe how predictors were handled in the analyses (functional form, rescaling, transformation, or any standardization).	44-45
	12c	D	Specify the type of model, rationale ² , all model-building steps, including any hyperparameter tuning, and method for internal validation	46-48
	12d	D;E	Describe if and how any heterogeneity in estimates of model parameter values and model performance was handled and quantified across clusters (e.g., hospitals, countries). See TRIPOD-Cluster for additional considerations ³	NA
	12e	D;E	Specify all measures and plots used (and their rationale) to evaluate model performance (e.g., discrimination, calibration, clinical utility) and, if relevant, to compare multiple models	46-48
	12f	E	Describe any model updating (e.g., recalibration) arising from the model evaluation, either overall or for particular sociodemographic groups or settings	NA
	12g	E	For model evaluation, describe how the model predictions were calculated (e.g., formula, code, object, application programming interface)	NA
Class imbalance	13	D;E	If class imbalance methods were used, state why and how this was done, and any subsequent methods to recalibrate the model or the model predictions	44-45
Fairness	14	D;E	Describe any approaches that were used to address model fairness and their rationale	NA
Model output	15	D	Specify the output of the prediction model (e.g., probabilities, classification). Provide details and rationale for any classification and how the thresholds were identified	44
Training versus evaluation	16	D;E	Identify any differences between the development and evaluation data in healthcare setting, eligibility criteria, outcome, and predictors	NA
Ethical approval	17	D;E	Name the institutional research board or ethics committee that approved the study and describe the participant-informed consent or the ethics committee waiver of informed consent	52-53
OPEN SCIENCE				
Funding	18a	D;E	Give the source of funding and the role of the funders for the present study	NA
Conflicts of interest	18b	D;E	Declare any conflicts of interest and financial disclosures for all authors	NA
Protocol	18c	D;E	Indicate where the study protocol can be accessed or state that a protocol was not prepared	NA
Registration	18d	D;E	Provide registration information for the study, including register name and registration number, or state that the study was not registered	NA

Data sharing	18e	D;E	Provide details of the availability of the study data	NA
Code sharing	18f	D;E	Provide details of the availability of the analytical code ⁴	NA
PATIENT & PUBLIC INVOLVEMENT				
Patient & Public Involvement	19	D;E	Provide details of any patient and public involvement during the design, conduct, reporting, interpretation, or dissemination of the study or state no involvement.	NA
RESULTS				
Participants	20a	D;E	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	40,55
	20b	D;E	Report the characteristics overall and, where applicable, for each data source or setting, including the key dates, key predictors (including demographics), treatments received, sample size, number of outcome events, follow-up time, and amount of missing data. A table may be helpful. Report any differences across key demographic groups.	55-59,60,61
	20c	E	For model evaluation, show a comparison with the development data of the distribution of important predictors (demographics, predictors, and outcome).	NA
Model development	21		Specify the number of participants and outcome events in each analysis (e.g., for model development, hyperparameter tuning, model evaluation)	59-65
Model specification	22	D;E	Provide details of the full prediction model (e.g., formula, code, object, application programming interface) to allow predictions in new individuals and to enable third-party evaluation and implementation, including any restrictions to access or re-use (e.g., freely available, proprietary)	59-65
Model performance	23a	D;E	Report model performance estimates with confidence intervals, including for any key subgroups (e.g., sociodemographic). Consider plots to aid presentation.	65-68
	23b	D;E	If examined, report results of any heterogeneity in model performance across clusters. See TRIPOD Cluster for additional details. ³	NA
Model updating	24	E	Report the results from any model updating, including the updated model and subsequent performance	NA
DISCUSSION				
Interpretation	25	D;E	Give an overall interpretation of the main results, including issues of fairness in the context of the objectives and previous studies	78-89
Limitations	26	D;E	Discuss any limitations of the study (such as a non-representative sample, sample size, overfitting, missing data) and their effects on any biases, statistical uncertainty, and generalizability	NA
Usability of the model in the context of current care	27a	D	Describe how poor quality or unavailable input data (e.g., predictor values) should be assessed and handled when implementing the prediction model	NA
	27b	D	Specify whether users will be required to interact in the handling of the input data or use of the model, and what level of expertise is required of users	NA
	27c	D;E	Discuss any next steps for future research, with a specific view to applicability and generalizability of the model	91-92

Nota N A: not applicable

¹ D=items relevant only to the development of a prediction model; E=items relating solely to the evaluation of a prediction model; D;E=items applicable to both the development and evaluation of a prediction model

² Separately for all model building approaches.

³TRIPOD-Cluster is a checklist of reporting recommendations for studies developing or validating models that explicitly account for clustering or explore heterogeneity in model performance (eg, at different hospitals or centres). Debray et al, BMJ 2023; 380: e071018 [DOI: 10.1136/bmj-2022-071018]

From: Collins GS, Moons KGM, Dhiman P, et al. BMJ 2024;385: e078378. doi:10.1136/bmj-2023-078378

Figura 13. Lista Verificación (Checklist) de TRIPOD-AI.

ModPred IA EnfCardiovasc_aps 1.0^{®1}

Reglas	Predictores (factores de riesgo)						Riesgo
r1	app hta/dm	colldl >4,16 mmol/L	pcr-as > 3 mg/dL	triglic >=2,3 mmol/L			Sí
r2	app hta/dm	colhdl <1,29/1,03 mmol/L	colldl >4,16 mmol/L	pcr-as > 3 mg/dL			Sí
r3	app hta/dm	colhdl <1,29/1,03 mmol/L	colldl >4,16 mmol/L	pcr-as 1 - 3 mg/dL	edad >= 58 años		Sí
r4	app hta/dm	colhdl <1,29/1,03 mmol/L	colldl >4,16 mmol/L	pcr-as 1 - 3 mg/dL	tas 140-159 mmHg	Fumar. Sí	Sí
r5	app hta/dm	pcr-as > 3 mg/dL	triglic >=2,3 mmol/L	tas 140-159 mmHg	edad <= 59 años		Sí
r6	app hta/dm	colhdl <1,29/1,03 mmol/L	pcr-as > 3 mg/dL	Fumar. Sí			Sí
r7	pcr-as > 3 mg/dL	coltot >=7,2 mmol/L)	edad >= 71 años				Sí
r8	pcr-as > 3 mg/dL	colldl >4,16 mmol/L	glicay = 5,6-7.0 mmol/L	edad >= 53 años			Sí
r9	app hta/dm	colhdl <1,29/1,03 mmol/L	pcr-as > 3 mg/dL	circabd > 102/88 cm			Sí

r10	app hta/dm	colhdl <1,29/1,03 mmol/L	pcr-as > 3 mg/dL	imc = Sobrepeso		Sí	
r11	app hta/dm	colhdl <1,29/1,03 mmol/L	colldl >4,16 mmol/L	pcr-as 1 - 3 mg/dL	tas >= 160 mmHg)	Sí	
r12	app hta/dm	colhdl <1,29/1,03 mmol/L	colldl >4,16 mmol/L	pcr-as 1 - 3 mg/dL	tas = 140-159 mmHg	circabd > 102/88 cm	Sí
r13	app hta/dm	colhdl <1,29/1,03 mmol/L	colldl 3,38- 4,15 mmol/L	glicay = 5,6-7.0 mmol/L	tas = 140-159 mmHg	Fumar. Sí	Sí
r14		colhdl <1,29/1,03 mmol/L	triglic >=2,3 mmol/L	imc = Sobrepeso	edad >= 68 años		Sí
r15	app hta/dm	colhdl <1,29/1,03 mmol/L	colldl 3,38- 4,15 mmol/L	triglic >=2,3 mmol/L	tas = 140-159 mmHg		Sí
r16	pcr-as > 3 mg/dL	triglic = 1,7- 2,2 mmol/L	edad >= 56 años				Sí

Figura 14. ModPred IA EnfCardiovasc_ aps 1.0

Recomendaciones.

¹ModPred IA EnfCardiovasc_ aps 1.0[®] es la primera versión del M-IA basado en reglas de decisión, que es lo más comprensible dentro de los modelos predictivos comprensibles, con un código de colores para resaltar los predictores más significativos, fácil de usar en la consulta del médico, no requiere gran aseguramiento logístico, más

allá de las investigaciones de los pacientes, las reglas de decisión generadas por los algoritmos JRip y J48 no se superponen, sí el paciente no cumple la condición de una regla se pasa a la siguiente, sí no cumple ninguna regla no se considera paciente en riesgo, solo hubo una intervención mínima del autor, que considera que su implantación en la práctica médica permitirá perfeccionarlo y medir su impacto clínico en la prevención y control de las enfermedades cardiovasculares, por lo que considera las siguientes recomendaciones para su uso:

1. En pacientes con solo 2 o más factores de riesgo en una regla y no se cumplen condiciones en otras reglas, debe valorarse según juicio clínico o interconsulta medicina interna o cardiología.
2. Si el paciente presenta factores de riesgo en una regla y no comprende la edad, se considera en riesgo, se da valor relativo a la edad, recordar que las enfermedades cardiovasculares se presentan cada día en pacientes más jóvenes.
3. En ausencia de antecedentes de hipertensión arterial, debe valorarse la cifras de presión arterial sistólica, pues a medida que la presión aumenta a partir de 115 mmHg, también lo hace el riesgo relativo de desarrollar aterosclerosis y enfermedad cardiovascular.
4. En ausencia de antecedentes de diabetes mellitus y se cumplen demás condiciones en una regla y, no se cumple otra regla de decisión, debe valorarse según juicio clínico o interconsultar con medicina interna o cardiología.
5. Si el paciente presenta cifras de triglicéridos elevadas y de colesterol HDL disminuido, se sugiere aplicar el índice aterogénico del plasma TG/cHDL, sí es mayor de 2 existe alta probabilidad de dislipemia aterogénica o metabólica y sospecha de aterosclerosis subclínica.
6. Sí evidencia glucemia alterada en ayunas, obesidad abdominal y las alteraciones lipídicas antes mencionadas se debe sospechar resistencia a la insulina, muy frecuente en el sobrepeso, el síndrome metabólico, la obesidad y la diabetes mellitus.

7. Una proteína c reactiva ultrasensible entre 1 y 3 mg/dL, sin otros factores de riesgo en la regla de decisión, se recomienda tratar intensivamente los demás factores de riesgo presentes en el paciente, pues esas cifras de pcr-as se consideran como pacientes con riesgo moderado de enfermedad cardiovascular, dado el rol relevante de la inflamación crónica o de bajo grado en las enfermedades cardiovasculares, así como, permite utilizar la proteína c reactiva ultrasensible como un blanco u objetivo terapéutico, o en su defecto la proteína c reactiva cualitativa, para el control de la inflamación crónica o de bajo grado y además puede ayudar al control del riesgo residual de origen lipídico.
8. Considerar que las enfermedades inflamatorias, inmunológicas, que son consideradas modificadores no tradicionales del riesgo de enfermedad cardiovascular, o infecciosas, cursan con niveles de proteína c reactiva mayores de 10 mg/dL lo que requiere de la valoración clínica del paciente, no obstante, la pcr también es un blanco terapéutico para controlar la inflamación propia de esas enfermedades.
9. El modelo fue diseñado para pacientes asintomáticos, entre 35 a 74 años, sin enfermedad cardiovascular establecida u otras condiciones de alto riesgo cardiovascular, es decir en prevención primaria, pero se recomienda su uso en edades más avanzadas, restándole el gran peso de la edad como factor de riesgo, donde el riesgo es mayor y su control tiene un impacto más significativo, también, en prevención secundaria, pues la aplicación del modelo después de un episodio cardiovascular puede ayudar a identificar los pacientes con mayor riesgo de morbilidad y mortalidad.

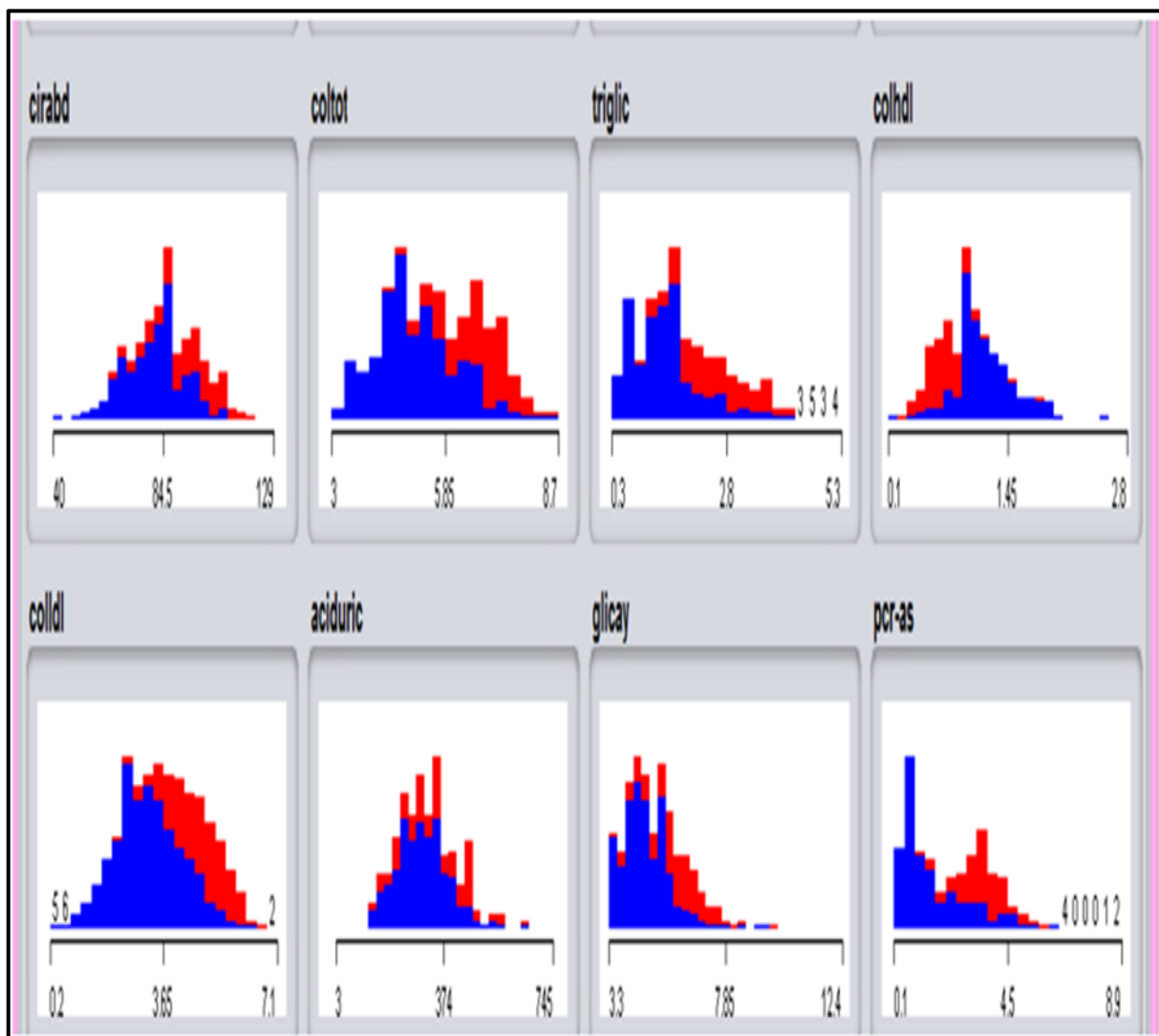


Figura 15. Rasgos predictores en interfaz gráfica 2D de Weka.

Tabla 9. Comparación de las características de la tabla de predicción de riesgo cardiovascular OMS 2019 con modelo predictivo implementado.

Características	<i>Tablas Predicción OMS 2019</i>	<i>ModPredIA EnfCardiovasc_ aps 1.0</i>
Tipo de estudio	Estimación de la carga de ECV ¹ para diferentes subregiones del mundo	Cohorte prospectiva
Nivel de evidencia de estudios pronósticos	II	I
Método empleado	Riesgos proporcionales de Cox	Inteligencia artificial
Validación	No	Sí
Estimación del riesgo a los 10 años	Porcentaje y categorías	Categorías
Categorías de riesgo	verde < 5 % amarillo 5 y <10 % naranja 10 y < 20% rojo 20 y < 30 % rojo fuerte > 30 %	Sí No
Edad de los participantes	40 a 79 años	35 a 74 años
Numero de variables	5	14
Tipo de variables	No modificables <ul style="list-style-type: none"> • sexo • edad Modificables <ul style="list-style-type: none"> • diabetes • tas • colesterol total 	No modificables <ul style="list-style-type: none"> • edad* Modificables <ul style="list-style-type: none"> • pcr-as* • circ. abdominal* • triglicéridos* • HTA (tas/tad) * • glucemias ayunas*

	<ul style="list-style-type: none"> • imc • hábito tabáquico 	<ul style="list-style-type: none"> • cHDL* • cLDL* • consumo tabaco* • DM 2^{2*} • IMC³ • colesterol total³
Aplicabilidad y facilidad de uso	Dos tablas de predicción de riesgo. Bajo costo especialmente sin colesterol en sangre	Tablas de predicción basado en reglas de decisión Variables clínicas y de laboratorio

¹ECV: Enfermedades cardiovasculares.

²La glucemia alterada en ayunas (prediabetes) fue un predictor más significativo, pero la DM2 es un factor de riesgo clásico, por si sola es un riesgo alto y los otros factores potencian ese riesgo.

³Se incluyeron factores tradicionales que son factores de riesgo establecidos en el inicio y progresión de las ECV.

*Predictores más significativos.

Tabla 10. Predicción del riesgo de enfermedad cardiovascular según tablas de riesgo.

ModPred IA EnfCardiovasc_ aps 1.0			Tablas Riesgo OMS 2019		
Riesgo	No	%	Riesgo	No	%
SÍ	519	31,8	Alto	387	23,6
			Intermedio	418	25,7
NO	1114	68,2	Bajo	828	50,7

Nota: n = 1633

Tabla 11. Predicción del riesgo de enfermedad cardiovascular según categorías de riesgo.

ModPred IA EnfCardiovasc_ aps 1.0			Tablas Riesgo OMS 2019		
Riesgo	No (n=1633)	%*	Riesgo	No (n=519)	%
Sí	519	31,8	alto	162	31,2
			intermedio	173	33,3
			bajo	184	35,5
Riesgo	No (n=1633)	%	Riesgo	No (n=1114)	%
NO	1114	68,2	alto	225	20,2
			intermedio	245	22,0
			bajo	644	57,8

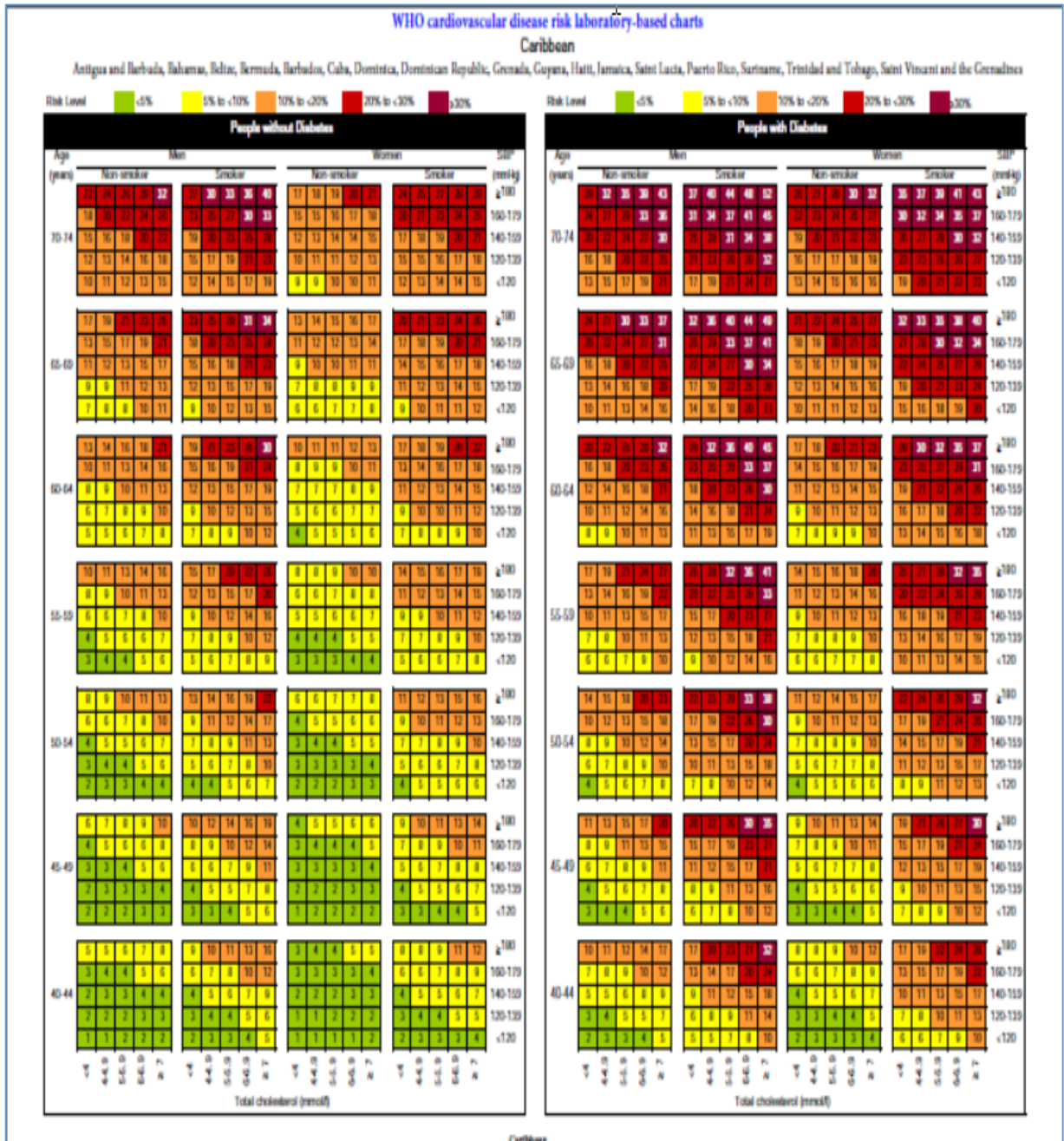


Figura 16. Tabla de predicción del riesgo cardiovascular de la OMS 2019 para Las Américas, región del Caribe.

