

---

This is the **published version** of the master thesis:

Arenas Villamizar, Ángel Ricardo; Suppi boldrito, Remo , dir. Algoritmo diagnóstico de Síndrome Coronario Agudo (SCA) basado en Inteligencia Artificial. 2024. 9 pag. (Màster en Intel·ligència Artificial i Big Data en Salut)

---

This version is available at <https://ddd.uab.cat/record/304354>

under the terms of the  license

# Algoritmo diagnóstico de Síndrome Coronario Agudo (SCA) basado en Inteligencia Artificial

## Angel Ricardo Arenas Villamizar

Resumen — Resumen del TFM:

El infarto de miocardio (IM) es la principal causa de muerte humana a nivel mundial. La electrocardiografía es la prueba de diagnóstico cardiovascular que se realiza con más frecuencia y la habilidad en la interpretación del electrocardiograma (ECG) es esencial para la mayoría de los médicos. Algunos estudios estiman la precisión diagnóstica entre estudiantes de medicina y especialistas entre un 55,8% y 74,9%. Se presenta un modelo de aprendizaje supervisado, de regresión logística y redes neuronales, a partir de la digitalización de electrocardiogramas físicos usando un software open-source que luego de procesar los datos con técnicas de normalización y ajustes, son etiquetados con el diagnóstico definitivo dado por coronariografía para ser probados en diferentes modelos. El objetivo principal de este estudio fue la generación de un algoritmo predictivo de síndrome coronario agudo a partir de electrocardiogramas de pacientes reales. Se tomaron datos de variables antropométricas de los pacientes, así como antecedentes cardígenicos y No cardígenicos que potencialmente afecten los trazados del electrocardiograma. Las bases de datos de los electrocardiogramas fueron agrupadas, normalizadas y consolidadas con los datos de cada paciente utilizando un ID. Los datos de identificación de los pacientes fueron anonimizados con un algoritmo de HASH y el protocolo de investigación fué presentado y aprobado por el Comité de Ética e Investigación de las clínicas de Red Salud en Santiago de Chile,

**Palabras clave** — Infarto de miocardio, inteligencia artificial, aprendizaje automático, electrocardiografía

### Abstract

Myocardial infarction (MI) is the leading cause of human death worldwide. Electrocardiography is the most frequently performed cardiovascular diagnostic test and skill in electrocardiogram (ECG) interpretation is essential for most physicians. Some studies estimate diagnostic accuracy among medical students and specialists between 55.8% and 74.9%. A supervised learning model of logistic regression and neural networks is presented, based on the digitization of physical electrocardiograms using open-source software that after processing the data with normalization techniques and adjustments, are labeled with the definitive diagnosis given by coronary angiography to be tested in different models. The main objective of this study was the generation of a predictive algorithm for acute coronary syndrome from electrocardiograms of real patients. The patients' antropometric variables were recorded, as well as cardiogenic and non-cardiogenic history potentially affecting the electrocardiogram tracings. The electrocardiogram databases were grouped, normalized and consolidated with the data of each patient using an ID. Patient identification data were anonymized with a HASH algorithm and the research protocol was submitted to and approved by the Ethics and Research Committee of the Red Salud clinics in Santiago, Chile.

**Index Terms**—Keywords: Myocardial Infarction, Artificial Intelligence, Machine Learning, Electrocardiography

## 1 INTRODUCCIÓN

El dolor torácico es la segunda razón más común por la que los adultos acuden a un departamento de emergencia en los Estados Unidos y representa más de 7 millones de visitas al año.<sup>1</sup>

El infarto de miocardio (IM) es uno de los eventos coronarios potencialmente mortales.<sup>2</sup> y es la presentación clínica más grave de la enfermedad de la arteria coronaria (EAC).<sup>3</sup>

De acuerdo con los cambios presentados en el electrocardiograma (ECG), esta patología se divide en dos categorías de IM con elevación del segmento ST (SCAEST) e IM sin elevación del segmento ST (SCASEST). Dado que la angina inestable es el antecedente inminente del IM, también se considera un estado de síndrome coronario

agudo (SCA).<sup>4</sup>

El infarto de miocardio (IM) es la principal causa de muerte humana a nivel mundial. Más de 3 millones de individuos desarrollan cada año un síndrome coronario con desviación del ST (SCAEST) (Figura 1) y más de 4 millones de personas presentan una patología de Síndrome coronario SIN desviación del ST (SCASEST).<sup>5</sup>



FIGURA 1. ELEVACIÓN SEGMENTO ST DERIVACIONES V1 Y V2

• E-mail de contacto: arenasvangelr@gmail.com  
 • Trabajo tutorizado por: Remo Suppi Boldrito <Remo.Suppi@uab.cat>  
 • Master IA aplicada en salud y BD.

En un metaanálisis se encontró que la prevalencia de infarto de miocardio en personas < 60 y > 60 años fue de 3,8% y 9,5%, respectivamente<sup>6</sup>

El diagnóstico de IM debe cumplir criterios tales como:

- Detección de aumento y/o disminución de biomarcadores cardíacos
- Síntomas de isquemia
- Cambios en el ECG indicativos de nueva isquemia (nuevos cambios en el segmento ST o nuevo bloqueo de rama izquierda del haz de His) (Figura 2)
- Desarrollo de ondas Q patológicas en el ECG
- Muerte súbita asociada a síntomas de isquemia, cambios en el ECG y/o evidencia de un trombo fresco en una angiografía coronaria y/o en la autopsia.<sup>7</sup>

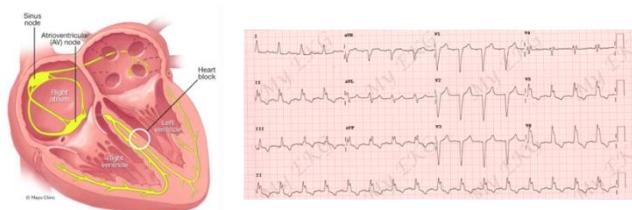


FIGURA 2. TOMADO Y MODIFICADO DE MAYOCLINIC – BLOQUEO DE RAMA Y MYECG

El músculo cardíaco se contrae por estímulos eléctricos generados en el nodo sinusal que se encuentra en la aurícula derecha y se transmiten por un sistema eléctrico al resto de las cavidades inferiores (ventrículos). En un bloqueo de rama, las vías por las que van estos impulsos se retrasan o se bloquean. La vía incluye dos ramas: las ramas izquierda y derecha. Si una rama está dañada, los ventrículos dependen de la otra rama para recibir y responder a las señales de la aurícula derecha. Si las dos ramas están bloqueadas, el corazón puede latir muy lentamente, lo que puede requerir un marcapasos; en un tercio de los pacientes, cuando esta situación ocurre de manera súbita, en el caso del bloqueo de rama izquierda, puede indicar el inicio de un IM. Esto se traduce en cambios en el trazado electrocardiográfico.

La electrocardiografía (ECG) convencional ofrece 12 imágenes (derivaciones) diferentes de la actividad eléctrica del corazón, representadas a partir de las diferencias de potencial eléctrico entre electrodos positivos y negativos colocados en los miembros y la pared torácica. Seis de estas derivaciones son verticales (emplean las derivaciones frontales I, II y III y las derivaciones de los miembros aVR, aVL y aVF) y 6 son horizontales (emplean las derivaciones precordiales V1, V2, V3, V4, V5 y V6).<sup>8</sup> (Figura 3)

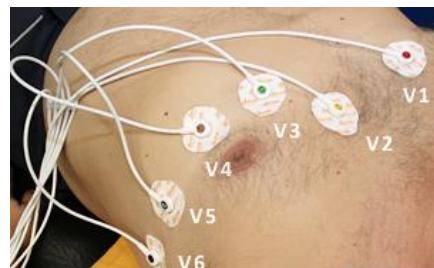


FIGURA 3. INSTALACIÓN DE ELECTRODOS UNIPOLARES V1 A V6 DE ECG. TOMADO DE [HTTPS://WWW.MADRID.ES/FICHEROS/SAMUR/DATA/603\\_01.HTM](https://WWW.MADRID.ES/FICHEROS/SAMUR/DATA/603_01.HTM)

El resultado final es producto de la superposición de los potenciales de acción de las células y músculos del corazón (Figura 4)

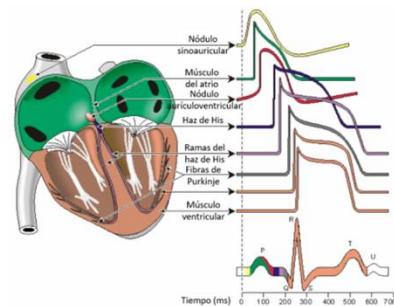


FIGURA 4. TOMADO Y MODIFICADO DE PACILIO, ENZO. DOI: 10.13140/RG.2.2.15558.42565

La señal electrocardiográfica registra la diferencia de potencial de voltaje que se genera en el corazón a través de electrodos colocados directamente sobre el paciente. La señal recogida por el electrocardiógrafo debe ser filtrada y amplificada. Los electrocardiógrafos digitales eliminan o suprimen las señales de baja frecuencia debidas a fluctuaciones por movimiento y respiración, así como las señales de alta frecuencia originadas por artefactos, movimientos musculares e interferencia electromagnética. Una vez que la señal ha sido filtrada y amplificada, se crea un complejo representativo para cada derivación a partir de la morfología dominante. (Figura 5)

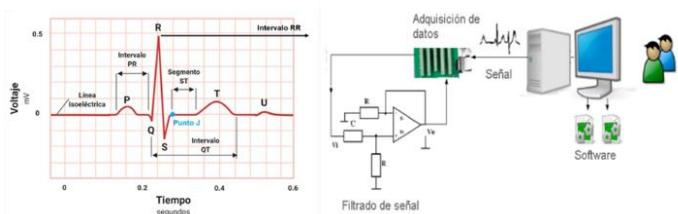


FIGURA 5. ELECTROCARDIOGRAMA. TOMADO DE [HTTPS://WWW.MAYOCLINIC.ORG](https://WWW.MAYOCLINIC.ORG)

Cook DA et al. publicaron un metaanálisis donde reunieron 78 estudios que evaluaron la precisión diagnóstica en la interpretación de ECG encontrando:

En estudiantes de Medicina, la precisión agrupada fue del 42,0% (IC del 95%, 34,3%-49,6%) mediana, 45%.

Para los residentes, la precisión agrupada fue del 55,8% (IC 95%, 48,1%-63,6%; mediana, 57%).

Para los médicos en ejercicio, la precisión agrupada fue del 68,5% (IC del 95%, 57,6%-79,5%; mediana, 66%).

Para cardiólogos y becarios de cardiología, la precisión agrupada fue del 74,9% (IC 95%, 63,2%-86,7%; mediana, 79%).<sup>9</sup>

La electrocardiografía, por lo tanto, es la prueba de diagnóstico cardiovascular no invasiva que se realiza con más frecuencia y la interpretación del electrocardiograma (ECG) es una habilidad esencial para la mayoría de los médicos.

Tradicionalmente los ECG son impresos y archivados dentro de las historias clínicas, en algunos lugares estos registros pueden ser digitales y existe una tendencia creciente para velar por esta práctica.

El uso de herramientas de IA viene demostrando una mejoría en la precisión diagnóstica en diferentes áreas de la Medicina.

El objetivo de este estudio fue diseñar un algoritmo para el diagnóstico de SCA basado en modelos de aprendizaje supervisado a partir de las imágenes de ECG de pacientes con diagnóstico de SCACEST y SCASEST confirmados mediante coronariografías.

## 2 METODOLOGÍA

### 2.1 Diseño

Se realizó un estudio retrospectivo, en dos clínicas de alta complejidad con servicios de urgencias, cardiología y hemodinamia en Santiago de Chile.

Los datos fueron procesados mediante la plataforma BigML utilizando algoritmos de aprendizaje supervisado.

El protocolo de investigación fue aprobado por el Comité de Ética e Investigación de Red Salud Chile.

### 2.2 Adquisición de datos

Se recolectaron los datos demográficos y clínicos generales: sexo, edad, talla, peso, tiempo de inicio de los síntomas, antecedentes cardiológicos: Hipertensión (HTA), Diabetes mellitus insulinorequiriente (DMIR), Diabetes mellitus no insulinorequiriente (DMNIR), tabaquismo, síndrome coronario previo con y sin instalación de stent y antecedentes no cardiológicos que pueden alterar el ECG: Tromboembolismo pulmonar (TEP), trastornos hidroelectrolíticos, hemorragia intracraneal, trauma craneoencefálico (TEC), procedimientos neuroquirúrgicos, meningitis, tumores intracraneanos, epilepsia, hipotermia, intoxicaciones con organofosforados, cianuro, metales pesados, monóxido de carbono, intoxicaciones con fármacos cardiotóxicos (digoxina, quimioterapia). Se consideró como

diagnóstico final el registrado en los protocolos de las coronariografías de cada paciente como Gold Standard de métodos invasivos.<sup>10, 11</sup>

Los datos demográficos, antecedentes y los diagnósticos de las coronariografías fueron digitados mediante un formato electrónico de registro de caso (CRF) y exportados a una hoja de cálculo. Cada paciente se identificó con un ID con el cual se identificó cada imagen de ECG.

Se descartaron los pacientes que no contaban con ECG previo a la CNG y pacientes con ECG de mala calidad

### 2.3 Transformación de datos

Se revisaron los archivos clínicos de dos clínicas de alta complejidad de Santiago de Chile de acuerdo a los diagnósticos dados por coronariografías (CNG) en los últimos 2 años. Se incluyeron aquellos ECG de pacientes que fueron tomados previos a la CNG y/o procedimientos de trombólisis farmacológica en urgencias, ya que los trazados electrocardiográficos cambian por el procedimiento y no solo por la progresión del IM.

Se descartaron los ECG de mala calidad, así como aquellos en los que no se contaba con el protocolo de las CNG.

Se revisaron más de 300 historias clínicas, desafortunadamente los ECG en papel son devueltos en su mayoría a los pacientes al momento del alta. Finalmente se tomaron 50 registros de electrocardiogramas (ECG) de 12 derivaciones, que cumplieron con los requisitos de ingreso al estudio.

Los ECG fueron procesados mediante **ECG-digitizing tool**, software *open source code*, que tiene la capacidad de transformar las imágenes de pixeles a señales digitales. En la validación del software la comparación muestra por muestra de señales registradas y digitalizadas mostró una correlación muy alta (0,977), una pequeña diferencia media (9,3  $\mu$ V) y un error cuadrático medio (25,9  $\mu$ V).<sup>12</sup>

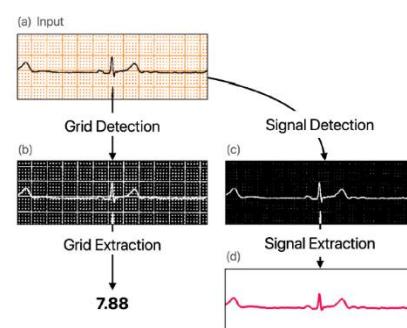


FIGURA 6. VISIÓN GENERAL Y EJEMPLO REPRESENTATIVO DEL PROCESO DE DIGITALIZACIÓN- TOMADO DE DOI.ORG/10.1016/J.CMPB.2022.106890

El software detecta inicialmente las escalas de grises de la imagen, posteriormente identifica las cuadrículas del ECG y define un umbral mediante el cual crea una nueva

imagen basado en ese umbral (Figura 6). El resultado es la digitalización de ECG con gran precisión. Dichas imágenes son exportadas a archivos csv.

Un total de 55 ECG fueron digitalizados utilizando **ECG-digitizing tool**, se descartaron 3 ECG por mala calidad de la imagen que no permitió la captura de la onda (Figura 7)



FIGURA 7. SEÑAL V4 QUE NO ES CAPTURADA POR MALA CALIDAD DEL TRAZADO.

Con las imágenes restantes (52) se capturaron las señales por cada derivación, una a la vez y se verificaron mediante la vista preliminar. (Figura 8)

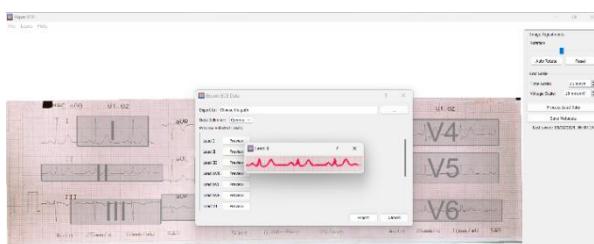


FIGURA 8. CAPTURA DE IMÁGENES POR CADA DERIVACIÓN.

El resultado final fueron 52 archivos csv con la información correspondiente a cada derivación, organizado en filas y columnas. (Figura 9)

FIGURA 9. DATOS ARCHIVO TXT IMPORTADOS EN EXCEL

## 2.4. Datos semisintéticos

Se generaron datos semisintéticos para compensar el tamaño de la muestra, utilizando técnicas estadísticas como SMOTE (sobremuestreo de minorías sintéticas), GAN (red generativa antagónica) y Cúpula Gaussiana obteniendo bases de 50, 100 y 10000 a partir de los datos originales. (Figura 10)

synthetic_data['data_Scaled'].head(10)								
1	2	3	4	5	6	7	8	9
_n_outliers	Vn_n_outliers	V5_n_outliers	V8_n_outliers	DI_n_outliers	AVI_n_outliers	DII_n_outliers	DIII_n_outliers	AVF_n_outliers
0.012282	0.025903	0.038465	0.024230	0.048748	0.034628	0.007215	0.018903	0.000000
0.012487	0.019795	0.026027	0.026760	0.031222	0.023439	0.025551	0.011312	0.000000
0.011192	0.008767	0.024904	0.044759	0.0275497	0.023749	0.034645	0.040937	0.000000
0.023062	0.034201	0.038360	0.002112	0.070206	0.025024	0.003663	0.019220	0.000000
0.124994	0.026531	0.026964	0.034742	0.131473	0.016371	0.034619	0.026935	0.000000
0.018259	0.023411	0.005084	0.016930	0.017422	0.008680	0.017159	0.011717	0.000000
0.016590	0.018994	0.022179	0.039985	0.031799	0.005634	0.073995	0.016486	0.000000
0.015136	0.019164	0.005196	0.020802	0.034264	0.014821	0.026060	0.028607	0.000000
0.005644	0.001218	0.000190	0.040208	0.041009	0.011230	0.002144	0.023567	0.000000

FIGURA 10. DATOS SEMI-SINTÉTICOS A PARTIR DE LOS 52 ORIGINALES.

## 2.5. Anonimización de datos

Los datos fueron pseudo-anonimizados usando la función de Hash que es un mecanismo que, aplicado a un dato concreto, genera una clave única o casi única que puede utilizarse para representar un dato. (Figura 11)



FIGURA 11. ANONIMIZACIÓN DE DATOS. FUNCIÓN DE HASH

## 2.6. Pre - Procesamiento de datos

Los datos generales de los pacientes y el diagnóstico de las CNG fueron exportados a una hoja de cálculo y se generaron 52 archivos *csv* luego del procesamiento en **ECG-digitizing tool**.



FIGURA 12. CONSOLIDACIÓN DE INFORMACIÓN POR PACIENTE

Los archivos con la información de los ECG fueron procesados utilizando Phyton en la plataforma Kaggle. Los registros de cada ECG se unieron a los datos de la hoja de cálculo utilizando el ID de cada paciente como clave primaria. (Figura 12)

Posteriormente se genera un nuevo dataframe con los datos consolidados, usando la función `cast` (Figura 13)

FIGURA 13. ARRAY CONJUNTO DE DATOS DE ECG

Los registros de las ondas fueron agrupados de acuerdo con la cara del corazón que representan (Figura 14):

Derivaciones derechas: V1-V2, ventrículo derecho y septo interventricular.

Derivaciones anteriores: V3-V4 pared anterior del

ventrículo izquierdo.

Derivaciones laterales bajas: V5-V6, pared lateral baja.

Derivaciones laterales altas: I y aVL, pared lateral alta.

Derivaciones inferiores: II, III y aVF, pared inferior

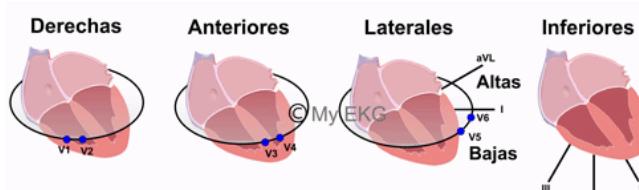


FIGURA 14. DERIVACIONES DERECHAS, ANTERIORES, LATERALES E INFERIORES DEL EKG. TOMADO DE [HTTPS://WWW.MY-EKG.COM/](https://www.my-ekg.com/)

Los datos producto de la digitalización de ECG fueron validados filtrando los datos inválidos con un umbral máximo del 5%. (Figura 15)

G1_f.head()								
shape: (5, 8)								
ID	V1	V2	Group	I	Invalid	Invalid Count	len	5%
12	99.958351	99.958351		0	0	1011	50.55	0
12	74.968763	99.958351		0	0	1011	50.55	0
12	74.968763	99.958351		0	0	1011	50.55	0
12	45.814244	99.958351		0	0	1011	50.55	0
12	16.659725	99.958351		0	0	1011	50.55	0

FIGURA 15. VALIDACIÓN POR GRUPOS

Dado que la distribución de los datos es diferente en cada derivada y que se encuentran datos extremos se ajustan los valores por encima y por debajo del 3% de los valores obtenidos.

A cada variable anterior se aplicó la función StandardScaled para eliminar la media y escalando a la varianza unitaria. La estandarización de un conjunto de datos es un requisito común para muchos estimadores de aprendizaje automático que pueden comportarse mal si las características individuales no se parecen más o menos a los datos distribuidos bajo una distribución normal (por ejemplo, gaussianos con media 0 y varianza unitaria).

Luego de realizar el escalamiento de los datos, unimos los grupos y consolidamos los datos el ID por paciente. (Figura 16)

data_clean_scaled1									
ID	V1_n_outliers	V2_n_outliers	V3_n_outliers	V4_n_outliers	V5_n_outliers	V6_n_outliers	D1_n_outliers	AVL_n_outliers	DII_n_outliers
12	164	164	164	164	164	164	164	164	164
47	0.035538	0.028628	0.03397	0.038217	0.004128	0.03508	0.016966	0.0	0.024293
37	0.031447	0.034591	0.042802	0.031128	0.027273	0.021018	0.027888	0.031873	0.028409
53	0.031967	0.020293	0.021018	0.047566	0.026256	0.023973	0.017582	0.01978	0.03408
33	0.044974	0.013228	0.027778	0.025253	0.037135	0.03183	0.039548	0.036723	0.008172
14	0.034247	0.034247	0.037383	0.034268	0.038339	0.044728	0.021605	0.018519	0.008688
...	...	...	...	...	...	...	...	...	...
24	0.035088	0.035088	0.021277	0.028369	0.037594	0.045113	0.035587	0.024911	0.03066
8	0.030303	0.029703	0.019544	0.035831	0.037855	0.028391	0.047445	0.056496	0.034314
50	0.031969	0.029412	0.028025	0.036946	0.018496	0.018496	0.019778	0.024722	0.023215
1	0.038961	0.038961	0.022113	0.031941	0.037681	0.034783	0.02611	0.028721	0.00216
6	0.034139	0.027027	0.02736	0.024624	0.026171	0.022039	0.02439	0.02439	0.003704

FIGURA 16. UNIÓN DE GRUPOS DESPUÉS DE ESCALAMIENTO

Una vez consolidadas las bases de datos se generaron los modelos de ML en BigML

### 3. PROCESAMIENTO DE DATOS

Se generaron 6 archivos de datos producto de cada procesamiento: 2 bases a partir de los datos reales con 52 registros y 4 bases con datos sintéticos con 50, 100, 500 y 10000 datos.

Para los datos producto de la digitalización de imágenes se utilizó PCA como técnica de reducción de dimensionalidad, se agruparon los datos a 5 grupos, eliminando los que podían producir ruido en la información.

La base de datos inicial contiene una muestra pequeña de 52 ECG, por lo que se consideró generar datos sintéticos.

Se probaron modelos de Random Forest, Regresión Logística, y Redes Neuronales tomando como variables objetivo: Enfermedad coronaria de 1, 2 o 3 vasos y arterias comprometidas agrupadas según la (o las) arterias reportadas en la coronariografía como lesiones Críticas (vaso culpable) y No críticas, comprometidas: ADA, ACx, ACD, ADA + TCI, ADA + ACx, ADA + ACD + ACx.

El mejor modelo de redes neuronales obtuvo un Accuracy del 88,89%, precisión 50%, recall 100%, coeficiente phi 0,66 mediciones que pueden mejorarse con el aumento del tamaño de la muestra.

Las herramientas de ML nos permiten modelar infinitud de escenarios clínicos

En nuestro caso, al modificar las variables en un modelo de aprendizaje supervisado por clasificación, se pueden estimar predicciones de interés clínico utilizando modelos de regresión logística y redes neuronales (Figura 17)

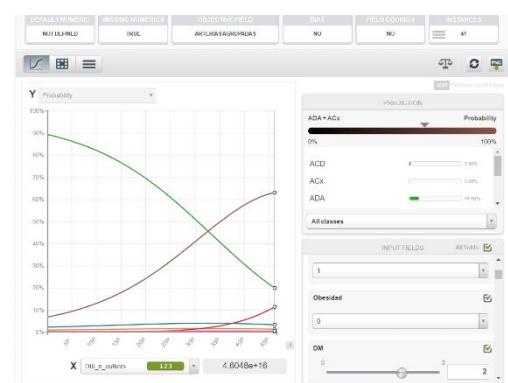


FIGURA 17. SEGÚN EL MODELO PARA LAS PACIENTES FEMENINAS CON DIABETES MELLITUS (DM) LA PROBABILIDAD DE DETECTAR UN COMPROMISO DE 2 O MÁS VASOS AUMENTA HASTA UN 62% CON LAS DEFLEXIONES EN DIII.



FIGURA 18. MODELO PREDICTIVO REGRESIÓN LOGÍSTICA

Para las pacientes femeninas con diabetes mellitus (DM) la probabilidad de compromiso de 2 o más vasos aumenta con las deflexiones en DIII. El modelo predice en un 70,99% la probabilidad de tener comprometida la ADA en mujeres con HTA y DMNIR. (Figura 18)

Esta probabilidad cambia al 60,75% de tener comprometida tanto la ADA como la ACx cuando se suman los antecedentes de tabaquismo y DMIR, lo cual concuerda con los resultados clínicos. (Figura 19)



FIGURA 19. MODELO PREDICTIVO REGRESIÓN LOGÍSTICA. AJUSTE DE VARIABLES

En otro escenario, en pacientes femeninas, sin diabetes con índice de masa corporal de 30 o más que las cataloga como obesas, el modelo sugiere que con las deflexiones de la derivación DI predominia el compromiso de la ADA llegando a un máximo de 40,01%, sin embargo al incluir el tabaquismo, predomina en esta derivada el compromiso de tres vasos ADA + ACD + ACx hasta un 41,57% (Figura 20)

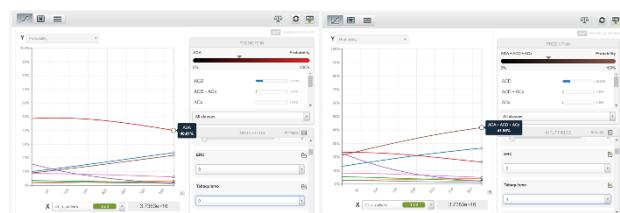


FIGURA 20. MODELO PREDICTIVO- REGRESIÓN MÚLTIPLE - MUJERES

En el caso de los hombres sin diabetes y obesos, el antecedente de tabaquismo también es relevante en los cambios electrocardiográficos en DI que pueden predecir el compromiso de tres vasos hasta un 41,48% (Figura 21)

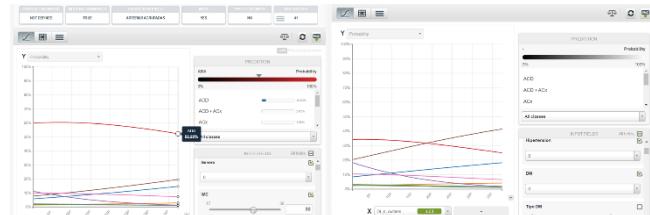


FIGURA 21. MODELO PREDICTIVO- REGRESIÓN MÚLTIPLE - MUJERES

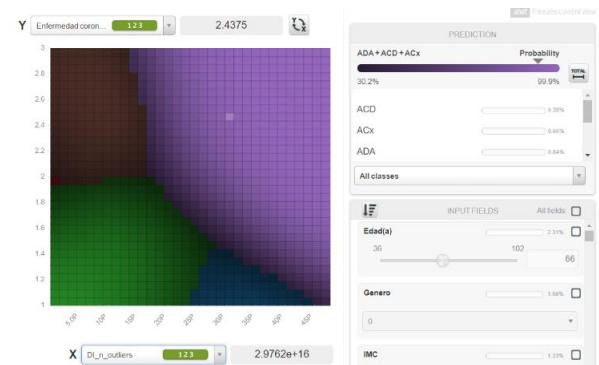
Utilizando un modelo de redes neuronales se estimó que las deflexiones en la derivación AVR (99,1%), V5 (96,9%), V1 (94,1%) son los mejores predictores de compromiso de tres vasos. Para la ADA los mejores predictores fueron V3 (97,5%), V1 (97,3%) y DII (96,4%) para ACD DI (77,8%), V3 (76,4%) y V2 (73,9%), para ADA + ACD AVF (79,9%), para ADA + ACx DIII(74,4%), para ADA + TCI V6 (48,2%). (Tabla 1)

	DI	DII	DIII	AVR	AVL	AVF	V1	V2	V3	V4	V5	V6
<b>ADA</b>	95,4	96,4	84,4	88,9	90,7	95,3	97,3	96,7	97,5	88,7	88,4	
<b>ACD</b>	77,8	43,2				54,3	46,1	73,9	76,4	65	67	35,3
<b>ACx</b>												
<b>ADA + ACD</b>	66	25,9	46,4	51	54,7	79,9	41,6		55,6	51,5	35,6	
<b>ADA + ACx</b>				74,4								57,4
<b>ADA + TCI</b>												48,2
<b>ADA + ACD + ACx</b>	84,4	90,5	73,2	99,1	73,2	90,8	94,1	91	72,2	72,5	96,9	77,3

TABLA 1. PREDICCIÓN CON REDES NEURONALES DE COMPROMISO SEGÚN VASO CORONARIO. ADA: Arteria descendente anterior, ACD: Arteria coronaria derecha, ACx: arteria circunfleja, TCI : Tronco coronario izquierdo

Las predicciones del modelo de redes neuronales se pueden visualizar mediante un mapa de calor por cada una de las derivadas electrocardiográficas- (Figura 22)

Figura 22. MODELO PREDICTIVO REDES NEURONALES. MAPA DE CALOR. DERIVACIÓN



DI.

### 3.1 Datos Sintéticos

Dada las características de la información, se eligen técnicas estadísticas con GAN, SMOTE, y Cúpula Gaussiana para generar archivos con 100, 1000 y 10000 datos semi-sintéticos a partir de los datos reales. Se evaluaron más de 200 modelos por cada uno, utilizando las herramientas de BigML de aprendizaje supervisado. (Figura 23) Para el caso de SMOTE la evaluación, utiliza K-means, pero ante una muestra tan pequeña, el sobremuestreo puede ser el responsable del overfitting. En el uso de GAN para generar los datos sintéticos, el mejor modelo, aunque logra un accuracy de 87,5%, su precisión fué del 27,27% y un recall de solo el 20%. Por lo tanto, no se obtuvo un adecuado desempeño de los modelos con los datos semisintéticos.

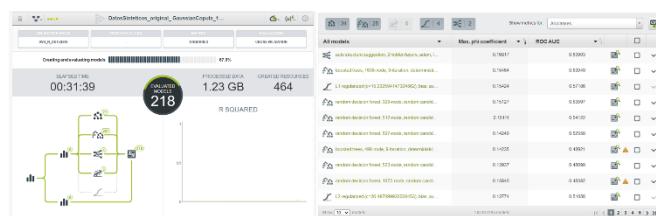


FIGURA 23. MODELOS GENERADOS A PARTIR DE DATOS SEMISINTÉTICOS POR MÉTODO DE CÓPULA GAUSSIANA

## 4. DISCUSIÓN

La electrocardiografía (ECG), mejorada mediante inteligencia artificial (IA), se ha convertido en una técnica potencial para el diagnóstico y tratamiento precisos de las enfermedades cardiovasculares<sup>13,14</sup> y permite diagnosticar patologías que antes no eran identificables mediante un electrocardiograma, o hacerlo con un rendimiento superior al que es hasta ahora posible<sup>15</sup>. Sin embargo, no se encuentra exento de limitaciones por la influencia de diferentes circunstancias como el género, raza, distribución poblacional, diferencias genéticas, comorbilidades, calidad de los equipos de ECG, políticas de archivo de registros clínicos, etc, que pueden generar sesgos en los datos de entrenamiento.

El futuro inmediato debe incluir la construcción de bases de datos robustas, confiables y con datos de alta calidad que permitan el adecuado procesamiento basados en herramientas de IA, aprovechando la generación a gran escala de datos de alta precisión de los equipos biomédicos que pueden conformar los Big Data clínicos.

En este estudio la principal limitante fue el acceso a los registros de ECG de los pacientes a su ingreso, ya que, al alta, gran parte de estos registros se entregan a los pacientes. Sumado a ello, la mala calidad de los ECG enviados desde la red extrahospitalaria limitó en algunos casos el adecuado reconocimiento de las ondas, por lo que fueron descartados.

Dado que la distribución de las ondas no tiene un comportamiento lineal ni sinusoidal, se consideró en el preprocesamiento, dividir el conjunto de datos en subconjuntos de validación, se utilizaron técnicas de normalización

(PCA) para reducir la dimensionalidad de los datos y técnicas de estandarización (StandarScaler). El análisis de componentes principales (PCA) es una técnica de reducción de dimensionalidad (número de características o columnas de entrada). La normalización escala cada variable de entrada por separado en el rango de 0 a 1, que es el rango para valores de punto flotante donde tenemos la mayor precisión. La estandarización escala cada variable de entrada por separado restando la media (lo que se denomina centrado) y dividiendo por la desviación estándar para cambiar la distribución para que tenga una media de cero y una desviación estándar de uno.<sup>16</sup> Sin embargo existe la posibilidad de que se sobreajuste al ser el conjunto de datos no representativo de la población en general.

Para los datos desbalanceados o los eventos de interés de baja frecuencia las técnicas de generación de datos sintéticos se convierten en una buena alternativa para el análisis, dando como resultado datos con mayor relevancia clínica en el caso de datos de salud, dada la multidimensionalidad de la enfermedad.

La creciente implementación del uso de herramientas de ML en la interpretación de datos clínicos, ineludiblemente traerán un beneficio de información completa y de calidad disponible para las atenciones hospitalarias y extramurales, en diferentes áreas de la Medicina

El crecimiento exponencial de herramientas basadas en IA y la falta de normativas, así como el creciente interés por la monetización de estas, hace necesario fortalecer la regulación y los mecanismos de validación que garanticen la mayor confiabilidad de los datos y la interpretación de los resultados, de forma tal que no impliquen un riesgo en la seguridad del paciente.<sup>17</sup>

## 5. CONCLUSIONES

En el presente trabajo se pretendía incluir más de 1000 ECG físicos ya que se contaba con una base de datos de dos años en tres clínicas diferentes; sin embargo, una vez revisadas las historias clínicas, se encontró que los ECG son entregados a los pacientes en su gran mayoría, sin que se guarde su registro físico o digital. Se revisaron más de 300 historias retrospectivamente sin encontrar los ECG, por lo que se dispuso una búsqueda prospectiva lo que afectó el tamaño muestral dado el corto tiempo disponible.

El trabajo por hacer es mantener esta recolección activa de información hasta lograr una muestra lo suficientemente grande con datos reales. Ya se dio el primer paso con el presente trabajo y se deja abierta la línea de investigación para trabajar en SCA con ML.

Se considera a mediano plazo el diseño de una app que facilite un acercamiento diagnóstico de SCA dirigida al personal clínico como herramienta de IA en su ejercicio.

una vez los modelos tengan un mejor rendimiento estadístico.

Para el procesamiento de imágenes de ECG el software **ECG-digitizing tool** tuvo un buen desempeño, sus limitantes principales radican en la calidad de los ECG.

El escalamiento de datos para este tipo de información es una buena alternativa para mejorar el rendimiento del modelo, en futuros trabajos se deben evaluar otras estrategias de normalización y agrupación de datos.

Para el análisis de las variables significativas se utilizó PCA, por cuestión de tiempo no se exploraron otras alternativas de agrupamiento, lo que se considera para futuros trabajos.

Las evaluaciones de los modelos con datos semisintéticos no fueron concluyentes en el presente estudio, sin embargo, es una buena elección cuando los datos son pocos o las frecuencias de los eventos demasiado bajas y generan datos desbalanceados. Se esperan mejores resultados en el futuro una vez se aumente el tamaño de la muestra de datos reales.

Los modelos de redes neuronales permiten predicciones con interés clínico en estudios con características similares al presente. Los resultados obtenidos en el presente estudio aún deben ser validados con trabajos con muestras más grandes.

## 6. AGRADECIMIENTOS

En primer lugar, a mis padres que ya hacen parte de las estrellas y que con su luz nos enseñaron a ver más allá de nuestros sueños.

A mi Familia y sobre todo a mis Hijos por ser el motivo que impulsa mis aventuras y que me inspiran para convertir lo imposible en tan solo una meta más por cumplir.

A mi Ana M, quien, con su fuerza, sus ganas de vivir y su Amor desde la distancia, me acompaña y hace de mis proyectos sus metas y me da la fuerza y la inspiración suficiente para lograrlas.

A Claudia Marcela Zapata Muñoz por su apoyo incondicional y motivación constante para lograr los objetivos. A Constanza Barroso Jotar por facilitar el acceso y la búsqueda de casos y su entusiasmo en este proyecto.

Agradecimientos a las Clínicas Red Salud Vitacura y Red Salud Providencia por facilitar el acceso a sus recursos para permitir este estudio.

## 7. REFERENCIAS

- [1] Center for Health Statistics N. National Hospital Ambulatory Medical Care Survey: 2015 Emergency Department Summary Tables. [http://www.cdc.gov/nchs/ahcd/ahcd\\_survey\\_instruments.htm#nhamcs](http://www.cdc.gov/nchs/ahcd/ahcd_survey_instruments.htm#nhamcs). Accessed May 1, 2018.
- [2] Hahla MS, Saeed Y, Razieh H. Comparison of risk factors & clinical and angiographic characterization of STEMI in young adults with older patients. *Res J Pharm Biol Chem Sci*. 2016;7(6):2013-2016.
- [3] Yeh RW, et al. Population trends in the incidence and outcomes of acute myocardial infarction. *N Engl J Med*. 2010;362(23):2155-2165. doi: 10.1056/NEJMoa0908610.
- [4] Thygesen K, et al. Universal definition of myocardial infarction. *J Am Coll Cardiol*. 2007;50(22):2173-2195. doi: 10.1016/j.jacc.2007.09.011.
- [5] Fox KAA, et al. Decline in rates of death and heart failure in acute coronary syndromes, 1999-2006. *J Am Med Assoc*. 2007;297(17):1892-1900. doi: 10.1001/jama.297.17.1892
- [6] Salari N, Morddarvanjoghi F, Abdolmaleki A, Rasoulooor S, Khaleghi AA, Hezarkhani LA, Shohaimi S, Mohammadi M. The global prevalence of myocardial infarction: a systematic review and meta-analysis. *BMC Cardiovasc Disord*. 2023 Apr 22;23(1):206. doi: 10.1186/s12872-023-03231-w. PMID: 37087452; PMCID: PMC10122825.
- [7] Thygesen K, Alpert JS, White HD; Joint ESC/ACCF/AHA/WHF Task Force for the Redefinition of Myocardial Infarction. Universal definition of myocardial infarction. *J Am Coll Cardiol*. 2007 Nov 27;50(22):2173-95. doi: 10.1016/j.jacc.2007.09.011. PMID: 18036459.
- [8] Cascino, Thomas. *Electrocardiografía Porta web Manual MSD*. Disponible en: <https://www.msdmanuals.com/es-cl/profesional/trastornos-cardiovasculares/pruebas-y-procedimientos-cardiovasculares/electrocardiograf%C3%ADA>
- [9] Cook DA, Oh SY, Pusic MV. Accuracy of Physicians' Electrocardiogram Interpretations: A Systematic Review and Meta-analysis. *JAMA Intern Med*. 2020 Nov 1;180(11):1461-1471. doi: 10.1001/jamainternmed.2020.3989. PMID: 32986084; PMCID: PMC7522782.
- [10] Orgullo YB, Tung P., Mohanavelu S., et al. Resultados angiográficos y clínicos entre pacientes con síndromes coronarios agudos que presentan depresión aislada del segmento ST anterior: un subestudio TRITON-TIMI 38. *J Am Coll Cardiol Cardiovasc Interv*. 2010; 3 (8): 806-811. doi: 10.1016/j.jcin.2010.05.012. [PubMed] [CrossRef] [Google Scholar]
- [11] Wang TY, Zhang M., Fu Y., Armstrong PW, Newby LK, Gibson CM, Moliterno DJ, Van de Werf F., White HD, Harrington RA, Roe MT Incidencia, distribución e impacto pronóstico de las arterias culpables ocluidas entre pacientes con síndromes coronarios agudos sin elevación del ST sometidos a angiografía diagnóstica. *Soy. Corazón J.* 2009; 157 (4): 716-723. doi: 10.1016/j.ahj.2009.01.004. [PubMed] [CrossRef] [Google Scholar]
- [12] J. D. Fortune, N. E. Coppa, K. T. Haq, H. Patel and L. G. Tereshchenko. Digitizing ECG image: a new method and open-source software code. *Computer Methods and Programs in Biomedicine* 2022 Pages 106890 DOI: <https://doi.org/10.1016/j.cmpb.2022.106890> <https://www.sciencedirect.com/science/article/pii/S0169260722002723>
- [13] Muzammil MA, Javid S, Afzadi AK, Siddineni R, Shahabi M, Haseeb M, Fariha FNU, Kumar S, Zaveri S, Nashwan AJ.

- Artificial intelligence-enhanced electrocardiography for accurate diagnosis and management of cardiovascular diseases. *J Electrocardiol.* 2024 Mar-Apr;83:30-40. doi: 10.1016/j.jelectrocard.2024.01.006. Epub 2024 Jan 28. PMID: 38301492.
- [14] Martínez-Sellés M, Marina-Breyssse M. Current and Future Use of Artificial Intelligence in Electrocardiography. *J Cardiovasc Dev Dis.* 2023 Apr 17;10(4):175. doi: 10.3390/jcdd10040175. PMID: 37103054; PMCID: PMC10145690.
- [15] Attia ZI, Harmon DM, Behr ER, Friedman PA. Application of artificial intelligence to the electrocardiogram. *Eur Heart J.* 2021 Dec 7;42(46):4717-4730. doi: 10.1093/eurheartj/ehab649. PMID: 34534279; PMCID: PMC8500024.
- [16] Jason Brownlee. How to Use StandardScaler and MinMaxScaler Transforms in Python. Disponible en: <https://machinelearningmastery.com/standardscaler-and-minmaxscaler-transforms-in-python/>
- [17] Ose B, Sattar Z, Gupta A, Toquica C, Harvey C, Noheria A. Artificial Intelligence Interpretation of the Electrocardiogram: A State-of-the-Art Review. *Curr Cardiol Rep.* 2024 Jun;26(6):561-580. doi: 10.1007/s11886-024-02062-1. Epub 2024 May 16. PMID: 38753291.