

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Departament d'Enginyeria Electrònica



Desarrollo de nuevos marcadores y clasificadores de bajo coste computacional para identificar afecciones cardiacas en registros ECG

Tesis Doctoral

Autor

Santiago Jiménez Serrano

Dirigida por

Dr. José Millet Roig

Dr. Francisco Sales Castells i Ramón

València, junio de 2023

Agradecimientos

Los agradecimientos de este trabajo suponen una mirada atrás de casi diez años, tiempo en el que he podido aprender y crecer tanto en lo personal como en lo profesional, gracias a multitud de personas que han compartido conmigo buenos y malos momentos y que se desviven en su día a día por dar lo mejor de sí mismas. Diez años dan para mucho, y en mi caso, no todo ha sido tesis, ya que la realidad de aprovechar las distintas oportunidades laborales que se presentan manda. En todo este tiempo, he tenido la oportunidad de trabajar en distintos centros de investigación, en la empresa privada, y también como profesor de instituto. Me gustaría hacer aquí mi pequeño homenaje a todas las personas de este camino que han sido importantes para mí, familia y amigos más cercanos incluidos.

En primer lugar, agradecer a mis directores de tesis: José Millet por haberme acogido en su grupo y mantener siempre una actitud positiva ante todo; y a Francisco Castells por haber tomado la iniciativa en la codirección de este trabajo, por haberme incluido en su proyecto de MCA, y por su incondicional apoyo y sabios consejos. Gracias también a Toni Guill y Álvaro Tormos por su enorme disposición a dotar de recursos el laboratorio y por esas jornadas en electrofisiología de la UV.

Gracias al equipo de Enfermedades Familiares del Hospital La Fe de València, especialmente a Esther Zorio, cuya calidad humana y profesional siempre me dejaba impresionado, y a Jorge Sanz, por su gran trabajo al generar y mantener la base de datos del estudio de MCA de esta tesis. Gracias por vuestra pasión en la investigación de nuevas soluciones y diagnósticos para enfermedades raras sin perder de vista la atención a vuestros pacientes.

También quiero agradecer a mis compañeros de laboratorio del grupo ITACA, en especial a Conra, Miguel, Alejandro, Ramón, Jorge, Ismael y Jaume, con los que he compartido grandes momentos: viajes a Pirineos, Madrid, despedidas y bodas, paellas en Xeraco, y aventuras varias con algunos de vosotros en Proteu. Mención especial para Edu, al que, aunque él no lo crea, admiro como persona, alguien en quien sabes que puedes confiar, con infinitas ganas de aprender y enseñar, y que, a pesar de las dificultades, se mantiene siempre de pie estóicamente.

Además, y aunque no tengan relación con esta tesis, quisiera agradecer a mis antiguos compañeros del IBV: a Iñaki Comas, por su gran labor durante la pandemia en el seguimiento del COVID, y por la gran flexibilidad, confianza y apoyo que me dio durante momentos difíciles; a Manoli, por hacerme ver que todavía queda gente

que antepone su pasión por la ciencia a un gran sueldo en una multinacional; y a Álvaro y Mariana por su incondicional y gran ayuda durante esa etapa.

También quiero agradecer a mis actuales compañeros y responsables del I3M y Oncovision, especialmente a Constan, que de alguna manera siempre me ha tenido apadrinado, y del que admiro esa capacidad para sacar tantas cosas adelante y poder acordarse de todo. También a José María Benlloch y a Antonio González por confiar en mí en esta nueva etapa que empiezo tras el doctorado. Y a Javi Cuallado porque sí, porque es un grande y se le quiere, aunque ya no nos veamos mucho.

Y ya para terminar este recorrido por los diez últimos años he de decir que, sin duda, ser profesor de instituto es lo que más me marcó en este periodo, dándome esa bofetada de realidad que creo, es necesaria para situarnos en el contexto real y social en el que nos encontramos. Mil gracias a los profesores y directores de centros públicos, que con los escasos recursos y apoyos que generalmente tienen, dedican su vida a enseñar a las nuevas generaciones no solo contenidos sino valores, independientemente de que el alumnado termine en la universidad o no. Y mil gracias a los y las estudiantes que tuve durante esos años, los cuales me enseñaron una enorme dignidad y entereza ante situaciones personales y sociales realmente complicadas, sin perder nunca esas ganas por lo realmente importante: disfrutar al máximo de lo que haces y lo que te gusta. Sin duda, todo esto no hizo más que reafirmarme en mi convicción de la necesidad de un sistema educativo público y de calidad que sirva para no dejar a nadie atrás, con recursos suficientes, profesorado motivado y motivador, y con visión estratégica tanto de comunidad como de país. Alumnos en los que me veía reflejado, ya que, sin la oportunidad de una escuela pública, aún con sus defectos, esta tesis, y tantas otras aventuras emprendidas, no pudieran haber tenido lugar.

Tampoco quiero olvidar a aquellos que me han acompañado más en el plano personal. A mis amigos que me quieren y me soportan, y se les quiere igual, sin ningún motivo más que por ser como son: Luis, Terry, Álex, Xavi, Gabal, David, Fer, Raúl. A la gente de Pachanguillas, porque cada partido es volver a sentirse un crío que salta la valla del colegio para simplemente correr detrás de un balón hasta morir, y porque muchas veces el tercer tiempo es el mejor momento de la semana. A Laura, por haberme apoyado siempre. Y a Ricar, porque desde esos viajes al Tour de Francia, o al norte en su Corsa desvencijado, hasta ahora, sabe que lo quiero como un hermano. Y a muchas otras personas que me dejo, pero que saben que deberían estar aquí.

También quiero agradecer a mi familia. A mis hermanos: Jesús, por su resolución en darnos a nuestro sobrino Samuel; a Antonio por su desparpajo y alegría; y a David por haber sido un apoyo y ayuda en todo momento, del que tengo mucho que aprender, y del que estoy más que orgulloso. A mi padre, Antonio, por haberme descubierto contra todo pronóstico en sus últimos años la Cantabria profunda, el sentido del perdón, y la dignidad ante la enfermedad. Y a mi madre, Virginia, por haber luchado siempre por sus cuatro hijos, y de la que, es más que

seguro, he heredado la tozudez y el empeño en todo lo que me propongo, además de haberme dado la educación, libertad, confianza y herramientas para poder ser quién soy ahora mismo. A Urbano por ser el mejor compañero que mi madre podía encontrar. A mis abuelos paternos, Manuel y Dolores, que con mucho sudor y esfuerzo llegaron de Andalucía a Valencia para mejorar sus vidas, y de los que estoy seguro, disfrutaron de sus nietos más que nadie. Y a mi abuela Rosa, con la que pasé los mejores veranos de mi infancia en La Solana.

Finalmente, quiero agradecer a María Luisa el haber compartido conmigo estos últimos casi cuatro años; por haberme acompañado en cualquier momento, bueno o malo, aceptándome tal cual soy. Por tener ese don para calmar y escuchar a la gente que lo necesita, y para ser la más alegre del lugar en cualquier fiesta; porque esos viajes y escapadas a cualquier sitio que parece el culo del mundo, contigo siempre son divertidos y únicos y no los cambiaría por ningún otro viaje. Y porque el día a día contigo es una aventura en la que siempre hay ganas de compartir, aprender, mejorar y disfrutar de la vida.

“Perquè hi haurà un dia que no podrem més i llavors ho podrem tot.”

V. A. Estellés

Abreviaturas

AA	Aprendizaje Automático
AS	Arritmia Sinusal
AUROC	Área bajo la curva ROC
AVN	Nodo Aurículoventricular – <i>Atrioventricular node</i>
BAVPG	Bloqueo Aurículoventricular de Primer Grado
BE	Eliminación hacia atrás – <i>Backward-Elimination</i>
BFAñI	Bloqueo Fascicular Anterior Izquierdo
biV	Biventricular
BR	Bloqueo de Rama
Bradi	Bradicardia
BradiS	Bradicardia Sinusal
BRD	Bloqueo de Rama Derecha
BRDC	Bloqueo de Rama Derecha Completo
BRDIn	Bloqueo de Rama Derecha Incompleto
BRI	Bloqueo de Rama Izquierda
BRIC	Bloqueo de Rama Izquierda Completo
CAP	Contracciones Auriculares Prematuras
CHi	Clasificadores Híbridos
CM	Métrica de la Competición – <i>Challenge Metric</i>
CNN	Red Neuronal Convolutiva – <i>Convolutional Neural Network</i>
CSu	Clasificadores Supervisados

CVP	Contracciones Ventriculares Prematuras
CTF	Criterios Diagnósticos <i>Task Force</i> 2010 para MCA
DEjD	Desviación del Eje a la Derecha
DEjI	Desviación del Eje a la Izquierda
DTW	Deformación Temporal Dinámica – <i>Dynamic Time Warping</i>
EC	Error Cuadrático
ECG	Electrocardiograma
ECHi	Entrenamiento de Clasificadores Híbridos
ECSu	Entrenamiento de Clasificadores Supervisados
ESP	Especificidad
FA	Fibrilación Auricular
FD	Frecuencia Dominante
FFNN	Redes neuronales <i>Feed Forward</i>
FLA	<i>Flutter</i> Auricular
FN	Falsos Negativos
FP	Falsos Positivos
FS	Selección hacia adelante – <i>Forward Selection</i>
GMM	Modelos de mezcla de Gaussianas – <i>Gaussian Mixture Model</i>
HMM	Modelos ocultos de Markov – <i>Hidden Markov Models</i>
HRV	Variabilidad de la Frecuencia Cardíaca – <i>Heart Rate Variability</i>
IA	Inteligencia Artificial
IC	Intervalo de Confianza
IIR	Respuesta Infinita al Impulso – <i>Infinite Impulse Response</i>
KDF	Función de densidad kernel – <i>Kernel Density Function</i>
KNN	K-Vecinos más cercanos – <i>K-Nearest Neighbors</i>
LOOCV	<i>Leave One Out Cross-Validation</i>
LSP	Latidos Supraventriculares Prematuros

LVP	Latidos Ventriculares Prematuros
MCA	Miocardiopatía Arritmogénica
MCA-biV	MCA con afectación biventricular
MCA-VD	MCA con afectación del ventrículo derecho
MCA-VI	MCA con afectación del ventrículo izquierdo
MPrR	Mala Progresión Onda R
MSC	Muerte Súbita Cardíaca
NaN	No es un número – <i>Not a Number</i>
NB	Naïve Bayes
OR	Otros Ritmos
PA	Potencial de Acción
PCA	Análisis de Componentes Principales <i>Principal Component Analysis</i>
PRPr	Intervalo PR Prolongado
QAn	Onda Q Anormal
QTc	Intervalo QT Corregido
QTPr	Intervalo QT Prolongado
RAmp	Amplitud media de la onda R
RBF	Función de Base Radial – <i>Radial Basis Function</i>
RD	Registro ruidoso
ReLU	Unidad Lineal Rectificada – <i>Rectified Linear Unit</i>
RFrag	Onda R Fragmentada
RL	Regresión Logística
RMp	Ritmo de Marcapasos
RMSE	Error Cuadrático Medio – <i>Root Mean Square Error</i>
RMSSD	Media cuadrática de diferencias sucesivas (entre intervalos RR adyacentes) – <i>Root Mean Square of the Successive Differences (between adjacent RR intervals)</i> .

RNA	Redes Neuronales Artificiales
RP	Reconocimiento de patrones
RSN	Ritmo Sinusal Normal
SAN	Nodo Sinusal – <i>Sinoatrial Node</i>
SEN	Sensibilidad
SVM	Máquinas de Vectores Soporte – <i>Support Vector Machines</i>
TacS	Taquicardia sinusal
TAn	Onda T Anormal
TArea	Área de la onda T
TCII	Trastorno de Conducción Intraventricular Inespecífico
TInv	Onda T Invertida
TNeg	Onda T Negativa
TV	Taquicardia Ventricular
TWAD	Dispersión del área de la onda T – <i>T-wave area dispersion</i>
VAI	Índice Angular Vectorial – <i>Vector Angular Index</i>
VCG	Vectocardiograma
VD	Ventrículo Derecho
VI	Ventrículo Izquierdo
VLI	Índice de Longitud Vectorial – <i>Vector Length Index</i>
VN	Verdaderos Negativos
VP	Verdaderos Positivos
VQRSB	Voltajes QRS Bajos
ZB	Selección de características por fuerza bruta
μ	Media
σ	Desviación estándar
Ku	Curtosis
Am	Asimetría

Contenido

Agradecimientos	ii
Abreviaturas	vi
Contenido	x
Lista de figuras	xvi
Lista de tablas	xviii
Resumen	xx
Resum	xxii
Abstract	xxiv
Capítulo 1. Introducción	1
1.1. Motivación	2
1.2. Hipótesis	4
1.3. Objetivos	4
1.4. Estructura de la tesis	5

Capítulo 2. Antecedentes y estado del arte7

2.1. Antecedentes Clínicos	7
2.1.1. Anatomía y electrofisiología del corazón	7
2.1.2. Ritmo sinusal normal	11
2.1.3. Fibrilación auricular	12
2.1.4. Otras arritmias cardíacas	13
2.1.5. Miocardiopatía Arritmogénica	16
2.2. Estado del arte	20
2.2.1. Marcadores de arritmias basados en la actividad auricular	20
2.2.2. Marcadores de arritmias basados en la actividad ventricular	22
2.2.3. Aprendizaje automático para el diagnóstico de arritmias	25
2.2.4. Detección de arritmias en dispositivos móviles	28

Capítulo 3. Materiales y métodos generales 31

3.1. Aprendizaje automático.....	31
3.1.1. Redes Neuronales Artificiales.....	34
3.1.2. Naïve Bayes	39
3.1.3. Support Vector Machines.....	40
3.1.4. Regresión logística.....	43
3.2. Sensibilidad y Especificidad	45

Capítulo 4. Detección de Fibrilación Auricular en una derivación única 47

4.1. Introducción.....	47
4.1.1. Objetivos.....	47
4.2. Materiales.....	48
4.3. Métodos	50
4.3.1. Criterios de evaluación	51
4.3.2. Preprocesado de señal	53
4.3.3. Extracción de características	53
4.3.4. Preprocesado del conjunto de datos	58
4.3.5. Selección de características.....	59
4.3.6. Entrenamiento y validación de Redes Neuronales	60
4.3.7. Entrenamiento y validación de Naïve Bayes.....	61

4.3.8. Entrenamiento y validación de <i>Support Vector Machines</i>	62
4.4. Resultados.....	63
4.4.1. Análisis de las características basadas en los vectores <i>pNNx</i>	63
4.4.2. Selección de Características	66
4.4.3. Redes Neuronales	69
4.4.4. Naïve Bayes.....	71
4.4.5. <i>Support Vector Machines</i>	73
4.4.6. Visión global de los resultados	74
4.4.7. Costes Computacionales.....	76
4.5. Discusión.....	79
4.5.1. Características basadas en los vectores <i>pNNx</i> medios	79
4.5.2. Selección de características	79
4.5.3. Resultados de clasificación y costes computacionales	80
4.5.4. Comparación con otros trabajos	80
4.5.5. Limitaciones y trabajos futuros	84
4.5.6. Conclusiones	84
Capítulo 5. Detección de múltiples afecciones cardiacas en diversos conjuntos de derivaciones.....	87
5.1. Introducción	87
5.1.1. Objetivos	88
5.2. Materiales	88
5.3. Métodos.....	92
5.3.1. Esquema de Validación	92
5.3.2. Reglas de Puntuación	93
5.3.3. Procesado de Señal	94
5.3.4. Extracción de Características.....	95
5.3.5. Preprocesado del Conjunto de datos.....	98
5.3.6. Selección de Características	98
5.3.7. Estrategia de Clasificación Binaria.....	99
5.3.8. Enfoque de Clasificación <i>One-vs-Rest</i>	103
5.4. Resultados.....	105
5.4.1. Costes Computacionales.....	105
5.4.2. Análisis de la Selección de Características.....	107
5.4.3. Análisis de la Selección de Modelos	110
5.4.4. Análisis de las Puntuaciones de los Modelos	111

5.5. Discusión	118
5.5.1. Resultados con una única derivación	118
5.5.2. Mejoras respecto el trabajo reportado en <i>CinC 2021</i>	119
5.5.3. Comparación con otros trabajos	121
5.5.4. Limitaciones y trabajos futuros	122
5.5.5. Conclusiones	123

Capítulo 6. Marcadores y clasificadores para detectar Miocardiopatía Arritmogénica 125

6.1. Introducción.....	125
6.1.1. Objetivos	127
6.2. Materiales.....	129
6.3. Métodos	130
6.3.1. Preprocesado de la señal	130
6.3.2. Extracción de parámetros basados en el ECG.....	132
6.3.3. Extracción de parámetros basados en el Vectocardiograma	134
6.3.4. Extracción de parámetros basados en PCA.....	136
6.3.5. Resumen de parámetros y análisis estadístico.....	137
6.3.6. Modelos de Regresión Logística y selección de características	139
6.3.7. Generación del conjunto de datos para entrenamiento de HMM... ..	141
6.3.8. División del conjunto de datos para HMM	142
6.3.9. Entrenamiento de HMM con HTK	143
6.3.10. Criterios de evaluación para los HMM	147
6.4. Resultados	148
6.4.1. Análisis estadístico de todo el conjunto de datos	148
6.4.2. Análisis estadístico estratificado por sexo	155
6.4.3. Modelos de Regresión Logística y selección de características	159
6.4.4. Modelos ocultos de Markov	164
6.5. Discusión	166
6.5.1. Criterios diagnósticos de MCA en el ECG	167
6.5.2. Modelos de Regresión Logística	168
6.5.3. Modelos ocultos de Markov	169
6.5.4. Limitaciones y trabajos futuros	169
6.5.5. Conclusiones	170

Capítulo 7. Conclusiones	171
7.1. Conclusiones	171
7.1.1. Respecto los objetivos principales	171
7.1.2. Respecto los objetivos secundarios	173
7.2. Directrices para trabajos futuros	178
Capítulo 8. Contribuciones	179
8.1. Contribuciones principales de esta tesis.....	179
8.1.1. Artículos en revista.....	179
8.1.2. Congresos internacionales	179
8.1.3. Congresos nacionales	180
8.2. Contribuciones relacionadas con esta tesis.....	181
8.2.1. Artículos en revista.....	181
8.2.2. Congresos internacionales	181
8.2.3. Congresos nacionales	182
8.3. Proyectos de investigación	182
Bibliografía	183

Lista de figuras

Figura 2.1. Anatomía del corazón	9
Figura 2.2. Red de conducción eléctrica del corazón.	9
Figura 2.3. Potencial de acción de un cardiomiocito.....	10
Figura 2.4. Representación esquemática del corazón y el ECG	10
Figura 2.5. Detalle de las ondas P-QRS-T e intervalos en el ECG	11
Figura 2.6. Ejemplo de señal ECG con ritmo sinusal normal	11
Figura 2.7. Ejemplo de señal ECG con fibrilación auricular.....	12
Figura 2.8. Ejemplos de distintos ritmos cardiacos normales y patológicos	14
Figura 2.9. Ejemplo de señal ECG con <i>flutter</i> auricular	15
Figura 2.10. Derivaciones precordiales de un paciente afectado por MCA-VD	18
Figura 2.11. Comparación de señales ECG con ritmo sinusal normal y FA	21
Figura 2.12. Gráficos de Poincaré.....	23
Figura 3.1. Generación de un modelo para el reconocimiento de patrones.....	32
Figura 3.2. Esquema básico de un perceptrón con n entradas.	35
Figura 3.3. Ejemplo de arquitectura de una red neuronal <i>Feed Forward</i>	36
Figura 3.4. Funciones de activación utilizadas en la capa oculta.	37
Figura 3.5. Idea básica de hiperplano en SVM.	40
Figura 3.6. Ejemplo de hiperplanos y margen de separación en SVM.....	41
Figura 3.7. Transformación mediante una función Kernel.....	42
Figura 3.8. Función <i>logsig</i> utilizada en las regresiones logísticas.....	44
Figura 4.1. Dispositivo Kardia Mobile (AliveCor)	49
Figura 4.2. Ejemplos de ECG de la base de datos utilizada	50
Figura 4.3. Proceso de filtrado de valores aberrantes en las señales ECG.	53
Figura 4.4. Ejemplos de detección de complejos QRS.....	54
Figura 4.5. Vectores $pNNx$ medios de cada una de las clases	64
Figura 4.6. Media y desviación de los vectores $pNNx$ (RSN y FA).	64
Figura 4.7. Diagramas de caja de los valores de <i>dtwRSN</i>	65
Figura 4.8. Diagramas de caja de los valores de <i>ecRSN</i>	65
Figura 4.9. Resultados de clasificación con FFNN según nº de unidades.....	70
Figura 4.10. Resultados de clasificación con FFNN	70
Figura 4.11. Resultados de clasificación con NB	72
Figura 4.12. Resultados de clasificación con SVM.....	74
Figura 4.13. Resultados de clasificación de los tres mejores modelos	75

Figura 4.14. Resultados de clasificación según el modelo.....	76
Figura 4.15. Rendimiento de los modelos de clasificación (muestras/segundo).....	77
Figura 4.16. Rendimiento de los modelos de clasificación (ms/muestra).....	77
Figura 4.17. Resultados de los 67 participantes en el <i>PhysioNet/CinC 2017</i>	83
Figura 5.1. Diagrama del entrenamiento de un Clasificador Supervisado.....	100
Figura 5.2. Diagrama del entrenamiento de un Clasificador Híbrido.	101
Figura 5.3. Diagrama del enfoque de clasificación <i>One-vs-Rest</i> empleado	104
Figura 5.4. Tendencia del tiempo medio de entrenamiento y <i>test</i>	106
Figura 5.5. Rendimiento medio de cada multclasificador de ECG	106
Figura 5.6. Porcentaje de características seleccionadas en cada categoría.	109
Figura 5.7. Métrica <i>G</i> de cada clasificador binario, según derivaciones.....	114
Figura 5.8. Métrica <i>G</i> de cada clasificador binario (media).....	117
Figura 6.1. Edad de los participantes del estudio de MCA.....	129
Figura 6.2. Filtrado de ECG de una persona del grupo control	131
Figura 6.3. Filtrado de ECG de una persona afectada por MCA-biV	131
Figura 6.4. Filtrado de ECG de una persona afectada por MCA-VI.....	132
Figura 6.5. Esquema básico de un Modelo Oculto de Markov.	144
Figura 6.6. Variables significativas del conjunto de datos <i>Dt</i> (1).....	149
Figura 6.7. Detección de ondas R fragmentadas.....	150
Figura 6.8. Detección de ondas T negativas.	150
Figura 6.9. Ilustración de los indicadores <i>midLevel</i> y <i>difLevel</i> (control).....	151
Figura 6.10. Ilustración de los indicadores <i>midLevel</i> y <i>difLevel</i> (MCA)	151
Figura 6.11. Variables significativas del conjunto de datos <i>Dt</i> (2).....	152
Figura 6.12. Vistas de un VCG de un control.....	153
Figura 6.13. Vistas de un VCG de un afectado por MCA	153
Figura 6.14. VCG del patrón QRS de un control, y un afectado por MCA.	154
Figura 6.15. Variables significativas en el conjunto de datos <i>Dfem</i>	156
Figura 6.16. Variables significativas en el conjunto de datos <i>Dmas</i>	157
Figura 6.17. Valores de salida de la regresión logística <i>DtλFS</i>	161
Figura 6.18. Valores de salida de la regresión logística <i>DtλBE</i>	161
Figura 6.19. Valores de <i>G</i> de los HMM en el conjunto de <i>test</i>	165
Figura 6.20. Valores de precisión de los HMM en el conjunto de <i>test</i>	166

Lista de tablas

Tabla 3.1. Matriz de confusión utilizada para contar aciertos y errores.....	45
Tabla 4.1. Nombre, abreviatura, número y duración de los registros ECG.....	48
Tabla 4.2. Reglas de conteo de aciertos y errores entre las clases a etiquetar.	51
Tabla 4.3. Detalle de las siete categorías de características extraídas del ECG.	55
Tabla 4.4. Valores de las características basadas en los vectores $pNNx$ medios. ...	63
Tabla 4.5. Características seleccionadas.	67
Tabla 4.6. Valores de las características seleccionadas según su clase ($\mu \pm \sigma$).	68
Tabla 4.7. Mejores resultados con redes neuronales	69
Tabla 4.8. Mejores resultados con Naïve Bayes.....	72
Tabla 4.9. Mejores resultados con SVM.....	73
Tabla 4.10. Ranking de los mejores resultados obtenidos según el criterio $F1$	75
Tabla 4.11. Resultados de clasificación obtenidos según el tipo de modelo	76
Tabla 4.12. Comparación de nuestros resultados con otros trabajos.....	82
Tabla 5.1. Número de registros ECG de la base de datos utilizada.....	89
Tabla 5.2. Ritmos cardíacos a detectar.....	90
Tabla 5.3. Número y conjuntos de derivaciones ECG utilizadas	91
Tabla 5.4. Detalle de las nueve categorías de características extraídas del ECG. ..	96
Tabla 5.5. Análisis de la selección de características – Cjtos. de derivaciones....	108
Tabla 5.6. Análisis de la selección de características – Derivaciones.	108
Tabla 5.7. Análisis de la selección de características – Categorías.	109
Tabla 5.8. Análisis de la selección de modelos – Cjtos. de derivaciones.	110
Tabla 5.9. Análisis de la selección de modelos – Afecciones cardíacas.	111
Tabla 5.10. Resultados de clasificación en la métrica CM	113
Tabla 5.11. Resultados de clasificación – Conjuntos de derivaciones.	113
Tabla 5.12. Resultados de clasificación – Conjuntos de derivaciones y afección	115
Tabla 5.13. Comparación de nuestros resultados con otros trabajos.....	122
Tabla 6.1. Datos poblacionales de la base de datos de MCA.....	129
Tabla 6.2. Resumen de las características extraídas de cada registro ECG.	138
Tabla 6.3. Registros ECG asignados para entrenamiento y test de HMM.	142
Tabla 6.4. Variables estadísticamente significativas (Dt).	148

Tabla 6.5. Variables significativas para muestras del sexo femenino (<i>Dfem</i>).....	155
Tabla 6.6. Variables significativas para muestras del sexo masculino (<i>Dmas</i>)....	156
Tabla 6.7. Comparación de variables sign. en la estratificación por sexo	158
Tabla 6.8. Resultados de clasificación de los modelos de regresión logística.	159
Tabla 6.9. Umbrales de clasificación de las regresiones logísticas.....	160
Tabla 6.10. Comparación de las variables selecc. para la Reg. Logística.....	163
Tabla 6.11. Resultados de clasificación de los HMM.....	164
Tabla 6.12. Evolución y estudios realizados con la base de datos MCA.....	166

Resumen

Las enfermedades cardiovasculares son una de las principales causas de mortalidad y morbilidad en el mundo. Entre las arritmias más comunes en adultos destaca la Fibrilación Auricular (FA), presentando una tendencia de crecimiento muy significativa, sobre todo en población de edad avanzada o con trastornos de obesidad. En el otro extremo, nos encontramos con la Miocardiopatía Arritmogénica (MCA), considerada una enfermedad rara con una prevalencia de 1:2000-5000 pero con gran afectación entre familiares directos, causante de muerte súbita cardíaca (MSC), y con un diagnóstico clínico complicado. Más allá de la FA o la MCA, existe una amplia variedad de patologías derivadas de una disfunción en la activación y conducción eléctrica del corazón.

Para todas ellas, el electrocardiograma (ECG) continúa figurando como la primera y principal técnica de diagnóstico clínico, siendo una herramienta fundamental de cribado y detección de patologías relativamente económica y ampliamente accesible. Sin embargo, el diagnóstico preciso a partir de la interpretación del ECG requiere de médicos experimentados, siendo ésta una tarea que consume recursos, tiempo y que además está sujeta a la variabilidad entre observadores.

Respecto a las afecciones cardíacas más comunes, conseguir un diagnóstico de forma automática que sea fiable, utilizando tanto 12 como un número reducido o único de derivaciones, sigue presentándose como un desafío. Este aspecto cobra especial relevancia con el uso cada vez más extendido de dispositivos portátiles o *wearables*, los cuales están ganando un gran interés para la detección temprana y preventiva de enfermedades cardíacas, registrando normalmente un número reducido de derivaciones ECG. Dicho uso masivo les confiere un gran potencial para facilitar el cribado y seguimiento de distintas afecciones en una amplia variedad de escenarios, a pesar de registrar señales de peor calidad en comparación con equipos certificados para uso clínico. El principal reto con estos dispositivos es encontrar un equilibrio adecuado entre la sensibilidad y la especificidad en la detección de ritmos cardíacos susceptibles de ser patológicos. En consecuencia, es indispensable diseñar e implementar algoritmos precisos adecuados para dispositivos móviles o portátiles capaces de detectar distintas afecciones cardíacas en registros de ECG.

Respecto las afecciones cardíacas menos comunes como el caso de la MCA, es necesario incrementar la sensibilidad en la detección durante los cribados intra-

familiares realizados tras una MSC. Para ello, sería posible explorar biomarcadores propios a esta enfermedad obtenidos mediante técnicas de procesado de señales ECG, además de modelos de clasificación que hagan uso de ellos, contribuyendo así a reducir el número de casos de muerte súbita.

En base a lo descrito anteriormente, la presente tesis estudia las posibilidades de diagnóstico basadas en técnicas de aprendizaje y clasificación automática en dos escenarios principales. El primero aborda la detección de la FA, así como un amplio abanico de otras patologías cardíacas comunes, donde proponemos y validamos distintos modelos de clasificación de bajo consumo computacional. Todo esto, utilizando extensas bases de datos de acceso abierto, y haciendo énfasis en enfoques de derivación única, ya que son los más utilizados en dispositivos móviles e inteligentes. El segundo escenario se centra en la detección de MCA mediante las 12 derivaciones estándar del ECG, donde proponemos y validamos nuevos biomarcadores y modelos de clasificación que tratan de incrementar la sensibilidad de los cribados intra-familiares realizados tras una MSC. Para ello, utilizamos una base de datos específica de la Unidad de Cardiopatías Familiares del Hospital Universitario y Politécnico La Fe de València.

Resum

Les malalties cardiovasculars són una de les principals causes de mortalitat i morbiditat en el món. Entre les arrítmies més comunes en adults destaca la Fibril·lació Auricular (FA), presentant una tendència de creixement molt significativa, sobretot en població d'edat avançada o amb trastorns d'obesitat. En l'altre extrem, ens trobem amb la Miocardiopatia Arritmogènica (MCA), considerada una malaltia rara amb una prevalença de 1:2000-5000 però amb gran afectació entre familiars directes, causant de mort sobtada cardíaca (MSC), i amb un diagnòstic clínic complicat. Més enllà de la FA o la MCA, existeix una àmplia varietat de patologies derivades d'una disfunció en l'activació i conducció elèctrica del cor.

Per a totes elles, l'electrocardiograma (ECG) continua figurant com la primera i principal tècnica de diagnòstic clínic, sent una eina fonamental de cribatge i detecció de patologies relativament econòmica i àmpliament accessible. No obstant això, el diagnòstic precís a partir de la interpretació del ECG requereix de metges experimentats, sent aquesta una tasca que consumeix recursos, temps i que a més està subjecta a la variabilitat entre observadors.

Respecte a les afeccions cardíques més comunes, aconseguir un diagnòstic de manera automàtica que siga fiable, utilitzant tant 12 com un número reduït o únic de derivacions, continua presentant-se com un desafiament. Aquest aspecte cobra especial rellevància amb l'ús cada vegada més estès de dispositius portàtils o *wearables*, els quals estan guanyant un gran interès per a la detecció precoç i preventiva de malalties cardíques, registrant normalment un nombre reduït de derivacions ECG. Aquest ús massiu els confereix un gran potencial per a facilitar el cribatge i seguiment de diferents afeccions en una àmplia varietat d'escenaris, malgrat registrar senyals de pitjor qualitat en comparació amb equips certificats per a ús clínic. El principal repte amb aquests dispositius és trobar un equilibri adequat entre la sensibilitat i l'especificitat en la detecció de ritmes cardíacs susceptibles de ser patològics. En conseqüència, és indispensable dissenyar i implementar algorismes precisos adequats per a dispositius mòbils o portàtils capaços de detectar diferents afeccions cardíques en registres de ECG.

Respecte les afeccions cardíques menys comunes com el cas de la MCA, és necessari incrementar la sensibilitat en la detecció durant els cribatges intra-familiars realitzats després d'una MSC. Per a això, seria possible explorar

biomarcadors propis a aquesta malaltia obtinguts mitjançant tècniques de processament de senyals ECG, a més de models de classificació que facen ús d'ells, contribuint així a reduir el nombre de casos de mort sobtada.

Sobre la base del descrit anteriorment, la present tesi estudia les possibilitats de diagnòstic basades en tècniques d'aprenentatge i classificació automàtica en dos escenaris principals. El primer aborda la detecció de la FA, així com un ampli ventall d'altres patologies cardíques comunes, on proposem i validem diferents models de classificació de baix consum computacional. Tot això, utilitzant extenses bases de dades d'accés obert, i fent èmfasi en enfocaments de derivació única, ja que són els més utilitzats en dispositius mòbils i intel·ligents. El segon escenari se centra en la detecció de MCA mitjançant les 12 derivacions estàndard de l'ECG, on proposem i validem nous biomarcadors i models de classificació que tracten d'incrementar la sensibilitat dels cribratges intra-familiars realitzats després d'una MSC. Per a això, utilitzem una base de dades específica de la Unitat de Cardiopaties Familiars de l'Hospital Universitari i Politècnic La Fe de València.

Abstract

Cardiovascular diseases are one of the leading causes of mortality and morbidity worldwide. Atrial Fibrillation (AF) stands out among adults' most common arrhythmias, presenting a very significant growth trend, especially in the elderly population or those with obesity disorders. At the other extreme, we find Arrhythmogenic Cardiomyopathy (ACM), a rare disease with a prevalence of 1:2000-5000 but great affectation among direct relatives, causing sudden cardiac death (SCD), and with a complicated clinical diagnosis. Beyond AF or ACM, there is a wide variety of pathologies derived from dysfunctions in the activation or electrical conduction of the heart.

For all of them, the electrocardiogram (ECG) continues to appear as the first and foremost clinical diagnostic technique, being a fundamental tool for screening and detecting pathologies that is relatively cheap and widely accessible. However, accurate diagnosis based on ECG interpretation requires experienced physicians, as this task consumes resources, time and is subject to variability between observers.

For the most common cardiac conditions, achieving a reliable diagnosis automatically, using either 12 or a smaller or single number of leads, remains a challenge. This aspect is especially relevant with the increasingly widespread use of portable or wearable devices, which are gaining significant interest for the early and preventive detection of heart disease, typically recording a reduced number of ECG leads. Such massive use gives them great potential to facilitate screening and monitoring different conditions in different scenarios, despite registering signals of lower quality compared to equipment certified for clinical use. The main challenge with these devices is finding the right balance between sensitivity and specificity in detecting pathologic heart rhythms. Consequently, designing and implementing accurate algorithms suitable for mobile or portable devices capable of detecting different cardiac conditions in ECG recordings is essential.

Concerning less common cardiac conditions such as the case of ACM, it is necessary to increase the sensitivity in detection during intra-family screenings carried out after an SCD. Hence, it would be possible to explore specific biomarkers to this disease obtained through ECG signal processing techniques, as well as classification models that use them, thus contributing to reduce the number of cases of sudden death.

Based on the previously described, this thesis studies the diagnostic possibilities based on machine learning and classification techniques in two main scenarios. The first deals with detecting AF and a wide range of other common cardiac pathologies, where we propose and validate different classification models with low computational consumption. All this, using extensive open access databases, and emphasizing single-lead approaches, since they are the most used in mobile and smart devices. The second scenario focuses on detecting ACM using the standard 12-lead ECG, where we propose and validate new biomarkers and classification models that try to increase the sensitivity of intra-family screenings carried out after an MSC. For this task, we used a specific database of the Familial Cardiopathies Unit of the Hospital Universitario y Politécnico La Fe de València.

Capítulo 1

Introducción

Las enfermedades cardiovasculares son una de las principales causas de mortalidad y morbilidad en el mundo [Vaduganathan 2022, Virani 2021]. Estas enfermedades disminuyen la calidad de vida de los pacientes y ocasionan un coste muy elevado a los sistemas nacionales de salud [Kumar 2022, Birger 2021, Gheorghe 2018]. En este sentido, la importancia clínica de las arritmias cardíacas está aumentando, junto con su incidencia y prevalencia, estando estrechamente relacionadas con el envejecimiento de la población [Savarese 2022, Chow 2012].

Entre las arritmias más comunes en adultos destaca la Fibrilación Auricular (FA) [Hindricks 2021, Chug 2014], presentando una tendencia de crecimiento muy significativa, sobre todo en población de edad avanzada o con trastornos de obesidad [Wang 2004]. Además, en ocasiones la FA es difícil de diagnosticar debido a comportamientos paroxísticos y/o la posible ausencia de síntomas [Naccarelli 2009].

En el otro extremo, nos encontramos con la Miocardiopatía Arritmogénica (MCA), considerada una enfermedad rara de base genética [McKenna 2017], con un diagnóstico clínico complicado y una prevalencia de 1:2000-5000 [Corrado 2020, Calkins 2017, Corrado 2011, Sen 2008]. Sin embargo, al tener una transmisión dominante, la probabilidad de sufrir MCA aumenta drásticamente a 1:2 entre familiares directos afectados por esta cardiopatía [Sharma 2022]. Dicha enfermedad tiene un alto impacto social, económico y mediático pues ocasiona entre el 15 y el 25% de las muertes súbitas cardíacas (MSC) en menores de 35 años [D'Aliento 1995, Nava 1988], siendo la primera causa de MSC asociada al deporte en España e Italia [Corrado 2011, Suárez 2013].

En este sentido, tanto para enfermedades que afectan a gran parte de la población, como otras consideradas raras, el electrocardiograma (ECG) sigue siendo la técnica principal en el diagnóstico cardíaco [Kligfield 2002], siendo una herramienta de cribado y detección de patologías coste-eficiente y ampliamente accesible. Sin embargo, el proceso de interpretación del ECG requiere de médicos

experimentados, siendo ésta una tarea que consume recursos, tiempo y que además está sujeta a la variabilidad entre observadores [Bickerton 2019].

1.1. Motivación

Algunos trabajos previos han abordado el problema de la detección automática de anomalías en el ECG de diferentes maneras con el fin de ayudar en el diagnóstico de afecciones cardíacas. Los primeros enfoques se basaron principalmente en el análisis tiempo-frecuencia y sus características [Alexakis 2003, Chazal 2004, Christov 2006, Mahmoud 2006, Kostka 2007], o en las transformadas de señal de Fourier y *wavelets* [Minami 1999, Martínez 2004, Mahmoodabadi 2005, Yang 2013, Aqil 2015]. Posteriormente, se probaron técnicas de lógica difusa y aprendizaje automático para detectar enfermedades cardíacas [Vafaie 2014, Chen 2018]. En [Hagan 2021] se realizó un estudio comparativo de distintos métodos para la clasificación de cinco categorías de anomalías cardiovasculares, donde se probaron máquinas de vectores de soporte (SVM), redes neuronales artificiales y conjuntos de árboles de decisión, siendo éstos últimos los que mejor relación entre precisión y coste computacional presentaron. En [Zheng 2020a], el mejor rendimiento de clasificación entre cuatro categorías de ritmos cardíacos diferentes (ritmo sinusal, FA, bradicardia sinusal y taquicardias auriculares o ventriculares) se logró mediante un clasificador *Gradient Boosting Tree* utilizando registros de ECG de 12 derivaciones. En los últimos años, el aprendizaje profundo, basado principalmente en redes neuronales convolucionales (CNNs), también está cobrando importancia en las tareas de clasificación de segmentos de ECG, habiéndose utilizado para detectar la FA en una sola derivación [Xiong 2018, Rahimeh 2021, Krasteva 2021], o para mejorar la precisión en la anotación de las ondas P-QRS-T a través de un modelo CNN recurrente [Sodmann 2018]. Otros trabajos tuvieron como objetivo detectar hasta nueve afecciones cardíacas en grandes conjuntos de datos de ECG, en su mayoría utilizando CNNs [Dai 2021, Hua 2021, Jo 2021, Yang 2021, Zhang 2021]. Además, trabajos recientes han publicado resultados sobre la detección de un mayor número de anomalías cardíacas en bases de datos que contienen hasta 66.361 registros de 12 derivaciones utilizando diferentes *frameworks* de aprendizaje profundo [Giovanni 2021, Zhu 2021, Zhao 2022].

Todos los estudios anteriores se basan en cohortes de pacientes relativamente grandes ya que las enfermedades a identificar tienen una prevalencia global elevada, lo que propicia la existencia de distintas bases de datos con un gran número de estos registros ECG. No obstante, en el caso de las enfermedades raras, como la MCA, estos registros ECG son escasos y complicados de conseguir, no apareciendo en las bases de datos públicas, lo que dificulta las tareas de desarrollo y validación de modelos de clasificación. Es por ello que para estos casos, una mejor estrategia a seguir es el estudio detallado de los biomarcadores de cada enfermedad rara en concreto, y el uso de modelos de clasificación en casos especiales como el cribado

intra-familiar [Sharma 2022, Dedeoglu 2022, Militaru 2019, Sweet 2015, Quarta 2011].

En este sentido, en el caso de la MCA, se han identificado ciertas modificaciones morfológicas en el ECG indicativas de posible afectación. Entre otros, se han descrito inversiones en la onda T de las derivaciones precordiales derechas (V1-V3), así como parámetros relacionados con la duración y amplitud del QRS, la prolongación del QRS en V1, además de la observación de la onda *epsilon* en esta misma derivación [Corrado 2020, Corrado 2021, Sen 2008]. Sin embargo, otras técnicas que han demostrado su eficacia en el diagnóstico de distintas patologías de origen ventricular no han sido probadas en MCA. Por ejemplo, en [Pastore 2019] se revisó la utilidad del vectocardiograma (VCG) en la identificación con éxito de distintas patologías, como la hipertrofia ventricular izquierda y derecha. Por otra parte, [Bhattacharyya 2020] afirma que el VCG permite distinguir pacientes con disfunción ventricular de aquellos sanos. Otra técnica inexplorada en el diagnóstico de la MCA es el análisis de componentes principales (PCA) de la señal ECG, el cuál ha sido utilizado con éxito, por ejemplo, en la identificación de pacientes con bloqueo de rama izquierda [Feeny 2020], o en otras patologías de origen auricular [Castells2005a, Castells 2007, Bonizzi 2010, Bortolan 2022].

En base a lo descrito anteriormente y desde un punto de vista técnico, podría incrementarse la sensibilidad en la detección de la MCA durante los cribados intra-familiares realizados tras una muerte súbita, así como su grado de afectación, utilizando biomarcadores propios a esta enfermedad mediante la aplicación de distintas técnicas de procesamiento de señal al ECG, contribuyendo así a reducir el número de casos de MSC.

Finalmente, para el caso de afecciones cardíacas más comunes, sigue siendo un desafío el diagnóstico automático y preciso utilizando tanto 12 como un número reducido o único de derivaciones [Dunn 2018, Georgiou 2018]. En este sentido, los dispositivos portátiles o *wearables* están ganando un gran interés para la detección temprana y preventiva de enfermedades cardíacas, tanto en la investigación como en entornos clínicos, presentando normalmente la característica de registrar un número reducido de derivaciones ECG. Debido a esto, tienen un gran potencial para detectar y facilitar el seguimiento masivo de distintas afecciones en una amplia variedad de escenarios [Hermans 2022, Duncker 2021, Kumari 2017]. El principal reto en el diagnóstico con estos dispositivos es encontrar un equilibrio adecuado entre la sensibilidad y la especificidad en la detección de ritmos cardíacos susceptibles de ser patológicos. Por lo tanto, es indispensable diseñar e implementar algoritmos precisos y adecuados para su uso en dispositivos móviles o portátiles capaces de detectar distintas afecciones cardíacas en registros de ECG.

1.2. Hipótesis

La presente tesis sostiene dos hipótesis principales:

Hipótesis 1. Es posible generar modelos de clasificación de bajo consumo computacional adecuados para dispositivos móviles e inteligentes, que utilicen como entrada características extraídas de la señal ECG con el fin de detectar con una precisión adecuada distintas patologías cardíacas.

Hipótesis 2. Pueden definirse nuevos biomarcadores y clasificadores basados en el ECG mediante técnicas de procesamiento de señal para enfermedades raras como la MCA, que permitan incrementar la sensibilidad de los cribados intra-familiares realizados tras un episodio de muerte súbita cardíaca.

1.3. Objetivos

La presente tesis plantea dos objetivos principales:

Objetivo Principal 1. Desarrollar y validar metodologías robustas y de bajo coste computacional, capaces de identificar una amplia variedad de afecciones cardíacas a partir de distintas combinaciones de derivaciones ECG, haciendo énfasis en los escenarios de derivación única. Este objetivo pretende abordar la necesidad clínica de poder identificar automáticamente múltiples afecciones cardíacas usando nuevas tecnologías portátiles que trabajan con información electrocardiográfica limitada. También pretendemos evaluar los rendimientos de clasificación con segmentos cortos de ECG, ya que son los más utilizados en dispositivos móviles e inteligentes.

Objetivo Principal 2. Definir nuevos biomarcadores y clasificadores basados en el ECG para el análisis diagnóstico de la MCA que contribuyan a detectar nuevos casos durante cribados intra-familiares, ayudando así a reducir el número de muertes súbitas por esta enfermedad.

Para alcanzar los objetivos principales previos, se proponen los siguientes objetivos secundarios:

Objetivo Secundario 1. Desarrollar algoritmos para la extracción y selección automática de características temporales y espectrales derivadas de la actividad ventricular de las señales ECG.

Objetivo Secundario 2. Entrenar y validar el rendimiento de diversos modelos de clasificación que etiqueten registros cortos de ECG de una única derivación entre los siguientes cuatro grupos: ritmo normal, fibrilación auricular, otros ritmos y registros ruidosos.

Objetivo Secundario 3. Entrenar y validar el rendimiento de diversos modelos de clasificación que utilicen conjuntos de distintas derivaciones ECG, para etiquetar nuevas muestras entre 26 ritmos cardiacos diferentes, patológicos principalmente.

Objetivo Secundario 4. Definir y evaluar nuevos indicadores para el diagnóstico de MCA mediante registros ECG, analizando las diferencias existentes al estratificar por sexo.

Objetivo Secundario 5. Evaluar distintos modelos de clasificación, que identifiquen a afectados y no afectados por MCA utilizando registros cortos de ECG de 12 derivaciones, contemplando también la estratificación por sexo.

1.4. Estructura de la tesis

Esta tesis se estructura en siete capítulos principales junto a un octavo en el que se relacionan las contribuciones científicas derivadas. A continuación, se indica una breve descripción de cada capítulo, donde los capítulos 4-6 se corresponden con las publicaciones en revistas y congresos nacionales e internacionales realizadas durante la presente tesis.

Capítulo 1. Introducción. En este capítulo se expone la motivación para el desarrollo de este trabajo de investigación, las hipótesis y objetivos de la tesis, así como la estructura del documento.

Capítulo 2. Antecedentes y estado del arte. Consiste en una revisión de los antecedentes clínicos y el estado del arte en las diferentes disciplinas involucradas en este trabajo.

Capítulo 3. Materiales y métodos generales. El objetivo de este capítulo es describir los métodos comunes utilizados durante la tesis. Se describirá el marco teórico del aprendizaje automático, y el reconocimiento de patrones mediante distintos modelos de clasificación como las redes neuronales, SVM o Naïve Bayes. También se abordarán las definiciones básicas de sensibilidad y especificidad.

Capítulo 4. Detección de Fibrilación Auricular en una derivación única. Este capítulo presenta una metodología de bajo coste computacional capaz

de identificar la FA en registros ECG de una única derivación, evaluando los distintos rendimientos computacionales y de clasificación. Las señales utilizadas aquí provienen de un pequeño dispositivo comercial conectado a un teléfono móvil.

Capítulo 5. Detección de múltiples afecciones cardíacas en diversos conjuntos de derivaciones. En este capítulo se presenta una metodología de bajo coste computacional, capaz de identificar una amplia variedad de afecciones cardíacas a partir de seis combinaciones de derivaciones de ECG diferentes (12, 6, 4, 3, 2 y 1 derivación). En este caso se evalúa el rendimiento computacional y de clasificación de la solución propuesta con segmentos cortos de ECG.

Capítulo 6. Marcadores y clasificadores para detectar Miocardiopatía Arritmogénica. Este capítulo propone un conjunto de indicadores y modelos de clasificación con el fin de asistir al diagnóstico de la MCA en sus formas de afectación de ventrículo izquierdo (VI) y biventricular (biV).

Capítulo 7. Conclusiones. En este capítulo se enumeran las conclusiones y principales hallazgos presentados en esta tesis, y se proponen unas directrices para trabajos futuros.

Capítulo 8. Contribuciones. Este capítulo enumera las contribuciones científicas principales derivadas de esta tesis, así como aquellas asociadas a ella durante su realización. Además, se indican los proyectos de investigación con colaboración clínica en los que se ha participado.

Capítulo 2

Antecedentes y estado del arte

Este capítulo revisa los antecedentes de los diferentes campos de investigación que se discuten en esta tesis, que van desde la electrofisiología cardiaca, la interpretación electrocardiográfica o las distintas cardiopatías a evaluar. Asimismo, también revisa el estado del arte respecto las distintas técnicas utilizadas para detectar automáticamente dichas patologías en el ECG.

2.1. Antecedentes Clínicos

2.1.1. Anatomía y electrofisiología del corazón

La anatomía del corazón (ver Figura 2.1) se divide principalmente en cuatro cavidades: dos superiores llamadas aurícula izquierda y derecha, y dos inferiores denominadas ventrículo izquierdo y derecho, respectivamente. En el contexto del sistema circulatorio, el corazón actúa como una bomba electromecánica, donde la sangre llega a las aurículas, pasando a continuación a los ventrículos a través de las válvulas auriculoventriculares. A continuación, tras la contracción ventricular, la sangre es expulsada hacia los pulmones y los órganos periféricos respectivamente. En todas estas etapas, una actividad eléctrica coordinada es decisiva para una correcta contracción aurículo-ventricular.

En este sentido, las células musculares encargadas de la contracción del tejido cardiaco, denominadas cardiomiocitos, se encuentran tanto en las aurículas como en los ventrículos. En cada latido, los ventrículos soportan un mayor esfuerzo mecánico, por lo que concentran más cardiomiocitos que cualquier otra parte del corazón, y, en consecuencia, tienen más tejido muscular, como puede observarse en la Figura 2.2.

El potencial de acción (PA) es la señal eléctrica que se propaga a través de los cardiomiocitos, produciendo una respuesta mecánica contráctil. El estímulo inicial puede provenir de células vecinas que se encuentran en contacto, o por el sistema

nervioso autónomo. La Figura 2.3 muestra con más detalle la relación entre un PA en un cardiomiocito y su correspondiente respuesta mecánica de contracción.

Asimismo, existe una red que conduce y sincroniza los impulsos eléctricos desde la zona auricular a la ventricular a través del haz de His y las fibras de Purkinje. Esta propagación eléctrica, que sucede durante cada latido, produce cambios de voltaje que pueden registrarse mediante electrodos sobre la superficie corporal y originan la señal conocida comúnmente como electrocardiograma. Durante un latido o ciclo cardiaco, intervienen distintos elementos en la generación de un complejo P-QRS-T característico del ECG ilustrados en la Figura 2.4.

Cada latido es producido por la contracción sincronizada entre los cardiomiocitos auriculares y ventriculares. Existen dos marcapasos naturales de este sistema: el nodo sinusal (SAN) y el nodo aurículoventricular (AVN), siendo el primero el más relevante [Hume 2012]. Estos nodos estimulan el sistema de fibras de Purkinje que finalmente dispara los PA producidos en los cardiomiocitos ventriculares [Guyton 2006]. Estos PA ventriculares causan el complejo más destacado del ECG, denominado QRS, donde R indica el pico más elevado de la señal. Es por esto por lo que el intervalo RR se utiliza típicamente para indicar el ritmo cardiaco. Por otra parte, la onda P indica la activación auricular, mientras que la onda T indica la repolarización ventricular. La Figura 2.5 muestra en detalle tanto las ondas P-QRS-T, como los intervalos característicos PR, ST y QT, donde cada letra indica su punto de inicio o fin.

En resumen, todo lo mencionado anteriormente demuestra la estrecha relación entre una correcta propagación eléctrica en el corazón con un ritmo cardiaco adecuado. Son precisamente los PA, junto con el ritmo del nodo sinusal, la base de la propagación eléctrica en el corazón, donde el complejo P-QRS-T del ECG no es más que la suma de todas estas señales en distintos momentos y lugares de la anatomía del corazón.

Finalmente, para obtener las 12 derivaciones estándar del ECG, se colocan diez electrodos adhesivos en el paciente: cuatro en las extremidades y seis en el tórax. Las derivaciones de las extremidades (I, II, III, aVR, aVL y aVF) se obtienen mediante la colocación de los electrodos en los brazos y las piernas, mientras que las derivaciones precordiales (V1, V2, V3, V4, V5 y V6) se obtienen mediante la colocación de los electrodos en posiciones específicas del tórax. Cada derivación muestra la actividad eléctrica del corazón desde un ángulo o perspectiva única, lo que permite evaluar diferentes aspectos de su funcionamiento.

Juntas, las 12 derivaciones estándar permiten obtener una visión global de la actividad eléctrica del corazón, evaluar el ritmo cardiaco, detectar anomalías en la conducción eléctrica, así como valorar la presencia de lesiones. No obstante, la derivación I por separado también proporciona información valiosa sobre la actividad eléctrica del corazón, permitiendo evaluar la función cardiaca y detectar posibles anomalías en la conducción eléctrica al registrar la diferencia de potencial entre el brazo derecho y el izquierdo [Bayés 2014].

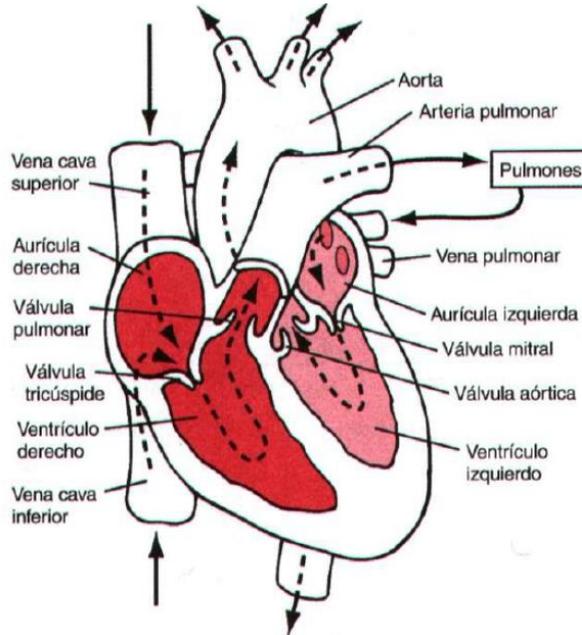


Figura 2.1. Anatomía del corazón [Guyton 2006].

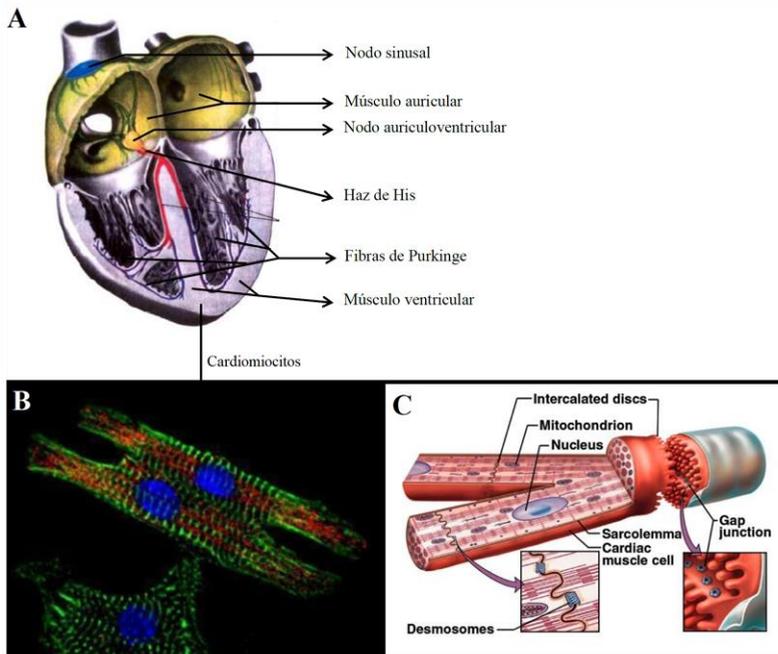


Figura 2.2. Red de conducción eléctrica del corazón (A). Imagen de inmunofluorescencia de cardiomiocitos en cultivo (B). Esquema de la estructura de los cardiomiocitos (C) [Saladin 2012].

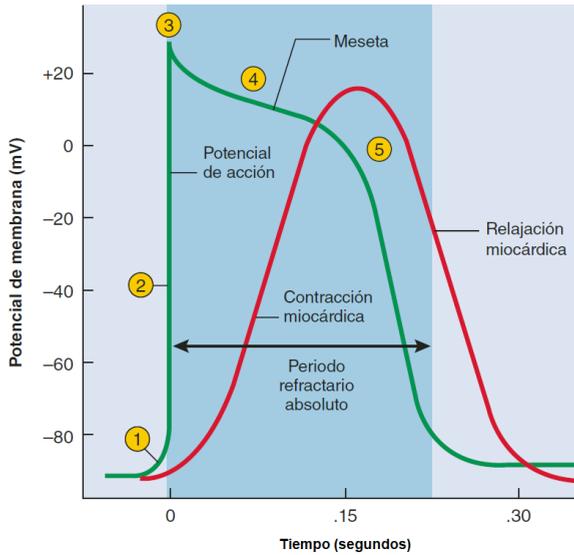
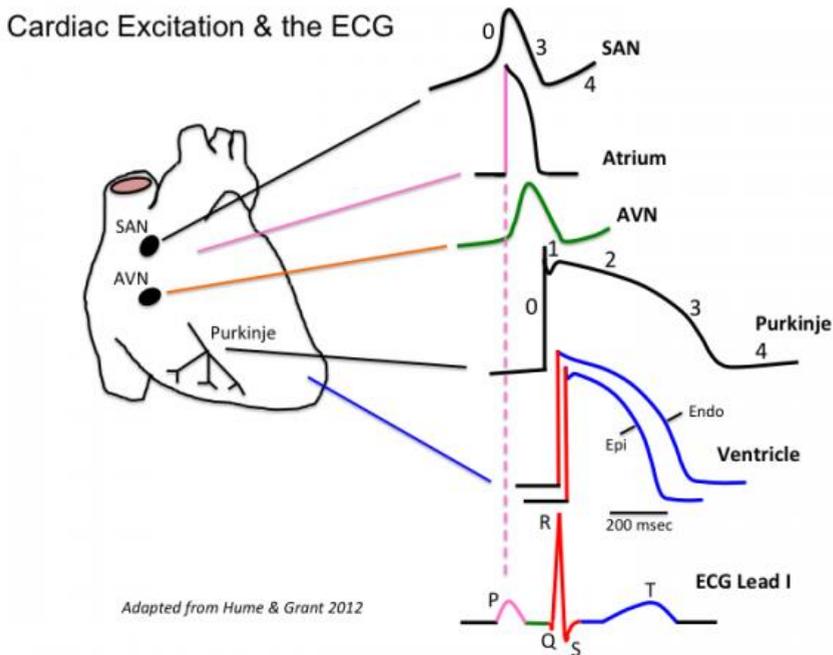


Figura 2.3. Potencial de acción (verde) de un cardiomiocito frente a su respuesta mecánica (rojo). [Saladin 2012].



Adapted from Hume & Grant 2012

Figura 2.4. Representación esquemática del corazón, el electrocardiograma (ECG), y los potenciales de acción (PA) registrados en diferentes regiones cardíacas. [Hume 2012].

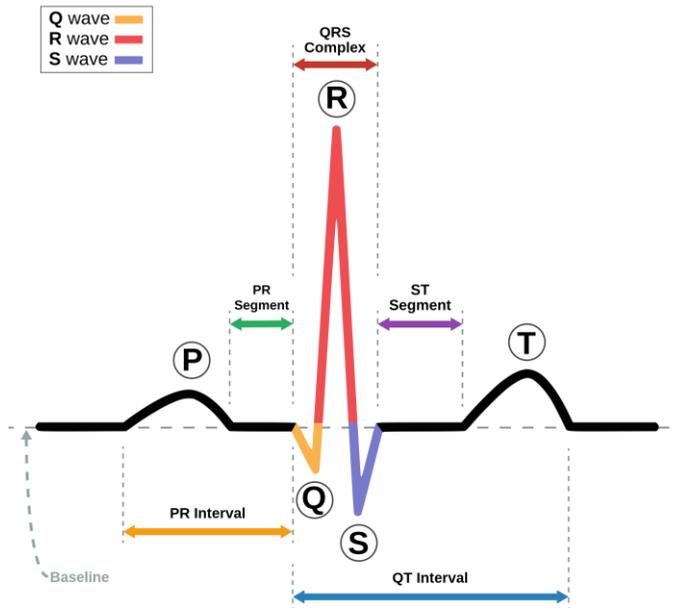


Figura 2.5. Detalle de las ondas P-QRS-T e intervalos que aparecen en el ECG [A. Atkielski, Wikimedia Commons, 2023].

2.1.2. Ritmo sinusal normal

Como su nombre indica, el ritmo sinusal normal (RSN) es el patrón rítmico del corazón que se origina en el nodo sinusal. Este ritmo cardiaco es el más común en individuos sanos, y su presencia en el ECG indica ausencia de patologías. El RSN se caracteriza por un ritmo entre latidos acompasado, con apenas variación temporal entre ondas R sucesivas, y con presencia de ondas P regulares y uniformes, indicando así una correcta activación auricular, tal como aparece en la Figura 2.6.

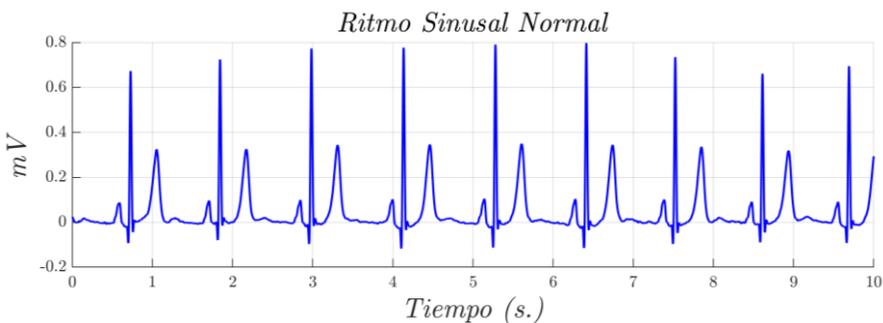


Figura 2.6. Ejemplo de señal ECG (derivación I) con ritmo sinusal normal (*Physikalisch-Technische Bundesanstalt database*).

Tras la onda P, se produce el complejo QRS, normalmente con una duración inferior a 0,12 segundos, indicando la contracción ventricular. Finalmente, después del complejo QRS, se presenta la onda T, que indica la repolarización ventricular.

El intervalo QT es otro parámetro relevante a evaluar en el ECG para determinar un RSN adecuado, representando el tiempo total necesario para la despolarización y repolarización ventricular. El intervalo QT se mide desde el inicio del complejo QRS hasta el final de la onda T, y su duración puede variar según la frecuencia cardíaca y otros factores como la edad, el sexo y la presencia de trastornos cardíacos o medicamentos que afecten a la conducción eléctrica del corazón.

En resumen, el RSN es un patrón rítmico caracterizado por la presencia de un intervalo RR regular, aunque modulado por el sistema nervioso autónomo, ondas P regulares, complejos QRS estrechos y una onda T adecuada a la correspondiente derivación del ECG.

2.1.3. Fibrilación auricular

La fibrilación auricular (FA) es la arritmia cardíaca sostenida más frecuente en población adulta, con una prevalencia global de hasta el 4%, y de hasta un 10% en mayores de 75 años [Hindricks 2021, Lip 2016, Wheeldon 1995]. Además, está relacionada con un riesgo significativo de mortalidad y morbilidad, y su prevalencia probablemente se triplicará en los próximos 30-50 años debido al envejecimiento demográfico en países occidentales [Mainardi 2008, Savelieva 2008].

La FA se define como una taquiarritmia caracterizada por una activación auricular predominantemente no coordinada, reflejada en el ECG por un ritmo RR irregular y ausencia de onda P, lo que provoca un deterioro de la función mecánica auricular [Fuster 2001]. Asimismo, en la FA aparece una línea base irregular en forma de ondas fibrilatorias en el intervalo TQ generalmente conocidas como ondas *f*. En la Figura 2.7 aparece una muestra de señal ECG correspondiente a FA donde se pueden apreciar todas las características descritas anteriormente.

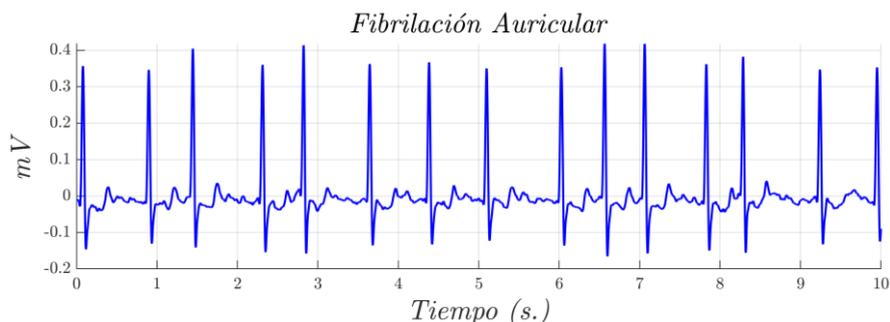


Figura 2.7. Ejemplo de señal ECG (derivación I) con fibrilación auricular (Physikalisch-Technische Bundesanstalt database).

En contraste con el RSN, donde un único frente de onda eléctrico se propaga desde las aurículas hacia los ventrículos, la FA presenta múltiples frentes de propagación caóticos y errantes en el corazón [Allessie 1995]. Dichos frentes de onda muestran frecuencias elevadas que predominan sobre el ritmo sinusal, lo que provoca contracciones inadecuadas. En casos extremos, esta frecuencia es tan elevada que los cardiomiocitos son incapaces de responder adecuadamente a todos los impulsos eléctricos recibidos, dando como resultado la actividad compleja, aparentemente desorganizada y aleatoria, denominada *fibrilación* [Jalife 2004].

La FA se manifiesta inicialmente en episodios cortos y/o esporádicos, aumentando progresivamente a episodios más largos y frecuentes. En este sentido, recientemente la clasificación de la FA ha variado [Hindricks 2021], debiendo estratificarse en la práctica clínica actual en las siguientes cinco categorías: diagnosticada por primera vez, FA paroxística, FA persistente, persistente de larga duración y permanente. Aquí, la FA paroxística implica episodios que terminan espontáneamente o con intervención en los 7 días siguientes al inicio, mientras que la FA persistente o persistente de larga duración implican duraciones superiores. Por último, la FA permanente se refiere a los casos en los que se decide no continuar con más intentos de restauración del RSN, aceptando la FA de forma perpetua e intentando controlar el ritmo mediante fármacos.

Entre las terapias para la terminación de la FA, se encuentra en primer lugar la administración de fármacos antiarrítmicos (cardioversión farmacológica) [Bollman 1998, Bollman 2002a, Bollman 2002b]. Si esta primera estrategia falla, se estudia la conveniencia de aplicar un choque eléctrico (cardioversión eléctrica), intentando así finalizar el proceso fibrilatorio al modificar o reiniciar las corrientes eléctricas caóticas del tejido miocárdico. Finalmente, se contemplan otras terapias más invasivas como la ablación de algunas zonas auriculares en un procedimiento quirúrgico. Además, se recomienda el uso de anticoagulantes en pacientes con FA, puesto que esta arritmia aumenta el riesgo de ictus debido a la formación de coágulos sanguíneos en el corazón que pueden viajar al cerebro y causar un accidente cerebrovascular [Hindricks 2021].

2.1.4. Otras arritmias cardiacas

Los distintos tipos de arritmias cardiacas se clasifican principalmente en función del origen y la frecuencia de la actividad eléctrica anormal en el corazón [Fu 2015, Gaztañaga 2012]. Respecto al origen anatómico de las arritmias, existen tres categorías principales: 1) arritmia auricular o supraventricular, 2) arritmia ventricular, y 3) arritmia de unión o aurículoventricular. Por otra parte, teniendo en cuenta el ritmo cardiaco, se habla de taquicardia al exceder los 100 latidos por minuto, mientras que el término bradicardia indica que no se superan los 60 latidos por minuto, ambos casos en población adulta. Además, la conducción eléctrica aurículoventricular puede ser normal, retardada o bloqueada. Finalmente,

dependiendo de la duración de la arritmia, se entiende que es no sostenida si su duración es inferior a 30 segundos, y sostenida en caso contrario.

El ECG muestra diferencias en las arritmias respecto un ritmo sinusal normal, ayudando a diagnosticar irregularidades del ritmo cardiaco, desviaciones en las rutas de conducción, agrandamiento o hipertrofia del músculo cardiaco, la presencia de un infarto de miocardio, o desequilibrios en los canales iónicos celulares, entre otros. La Figura 2.8 muestra algunos ejemplos al respecto de estas situaciones.

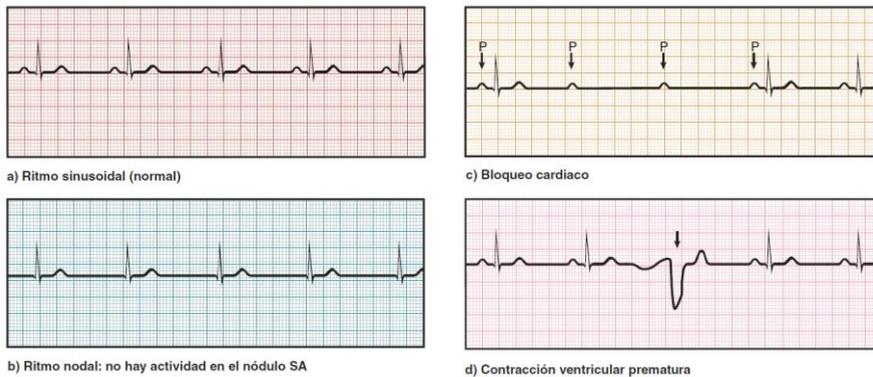


Figura 2.8. Ejemplos de distintos ritmos cardiacos normales y patológicos. [Saladin 2012].

A continuación, se enumeran y describen brevemente los diferentes tipos de arritmias cardiacas más comunes, excluyendo la FA ya descrita anteriormente.

Bradicardia sinusal: se produce cuando el nodo sinusal no genera impulsos eléctricos a una frecuencia adecuada, lo que provoca que el corazón lata más lentamente de lo normal. La bradicardia sinusal puede ser causada por enfermedades cardiacas, medicamentos o factores genéticos.

Ritmo nodal: es aquel generado por el nodo aurículoventricular en ausencia de actividad en el nodo sinusal, provocando ausencia de onda P (Figura 2.8, panel b).

Bloqueo cardiaco: se produce cuando la señal eléctrica que se transmite a través del sistema de conducción cardiaca se bloquea o retrasa, produciendo como resultado una frecuencia cardiaca lenta o irregular. Uno de los indicadores de bloqueo observables en el ECG es la ausencia de complejo QRS tras su correspondiente onda P (Figura 2.8, panel c).

Extrasístoles o contracciones ventriculares prematuras: son contracciones cardiacas que pueden ser percibidas como una palpitación o sacudida en el pecho (Figura 2.8, panel d). Aunque generalmente no son peligrosas, en algunos casos pueden indicar otra patología subyacente.

Taquicardia supraventricular: se origina en las aurículas o en el nodo sinusal, provocando una frecuencia cardiaca anormalmente rápida.

Flutter auricular: se produce cuando las aurículas se contraen a una frecuencia anormalmente alta, lo que puede provocar una frecuencia cardíaca rápida. Se caracteriza por ausencia de la onda P en el ECG, pero a diferencia de la FA, el *flutter* auricular presenta mayoritariamente un ritmo regular, como puede apreciarse en el ejemplo de la Figura 2.9.

Taquicardia ventricular: se origina en los ventrículos del corazón, provocando una frecuencia cardíaca anormalmente rápida y regular, y una forma inusual del complejo QRS. En casos graves, puede provocar un paro cardíaco al desencadenar una fibrilación ventricular, en la que desaparece cualquier coordinación mecánica para producir el correcto bombeo de sangre a los demás órganos, provocando así una muerte súbita cardíaca.

Los síntomas principales de las distintas arritmias suelen incluir palpitaciones, mareo, falta de aire, fatiga, y una frecuencia cardíaca demasiado rápida o lenta, normalmente irregular. Sin embargo, algunas arritmias presentan comportamientos asintomáticos, sobre todo en las primeras fases o estadios [Fu 2015].

Al igual que en la FA, tanto las arritmias supraventriculares como ventriculares pueden tratarse con la administración de fármacos antiarrítmicos o, en casos graves, mediante procedimientos quirúrgicos o ablativos con catéter. Respecto las arritmias ventriculares, los desfibriladores cardioversores implantables son la base del tratamiento de los pacientes con alto riesgo de muerte súbita ya que tienen capacidad para identificar y desfibrilar automáticamente en caso de arritmias ventricular malignas [Fu 2015]. No obstante, la desfibrilación se asocia a importantes efectos adversos físicos y emocionales. [Markman 2019].

En resumen, existen diferentes tipos de arritmias cardíacas que afectan a distintos aspectos del funcionamiento del corazón, comprometiendo su capacidad para bombear sangre de manera efectiva. Un diagnóstico temprano y adecuado es de especial relevancia, ya que permite detectar y prevenir complicaciones graves.

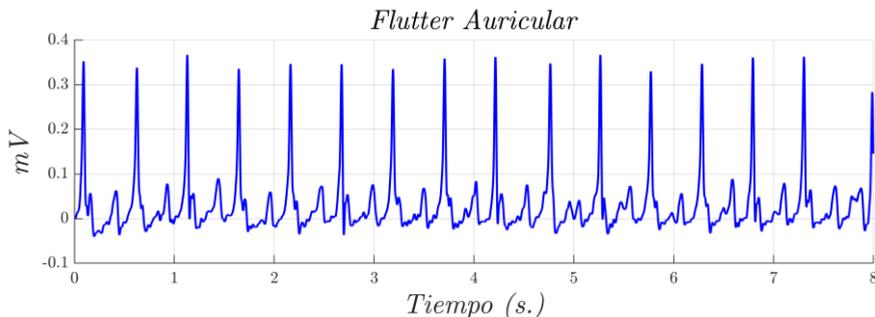


Figura 2.9. Ejemplo de señal ECG (derivación I) con *flutter* auricular (*Georgia 12-lead ECG Challenge database*).

2.1.5. Miocardiopatía Arritmogénica

La Miocardiopatía Arritmogénica (MCA) se define actualmente como una enfermedad genética del miocardio que afecta al Ventrículo Derecho (VD), al Ventrículo Izquierdo (VI) o a ambos (biventricular, biV), cuya característica más significativa es la sustitución miocárdica por una cicatriz fibrosa o fibroadiposa; esta cicatriz provoca disfunción ventricular global y/o local y predispone a desarrollar arritmias ventriculares potencialmente letales, independientemente de la gravedad de la disfunción ventricular sistólica [Sen 2008, Hoorntje 2017]. Recientemente, la MCA también se ha definido de forma más amplia en [Towbin 2019] de la siguiente forma: “trastorno arritmogénico del músculo cardíaco no explicado por enfermedad isquémica, hipertensiva o valvular”.

La MCA es considerada una enfermedad rara al presentar una prevalencia global de 1:2000 a 1:5000 en población general [Corrado 2011], además de presentar un origen hereditario [McKenna 2017]. En las últimas décadas ha tenido un gran impacto social y mediático por ser causa de muerte súbita cardíaca (MSC) en menores de 35 años en un 15–25% de los casos [Daliento 1995, Nava 1988]. Además, ocupa el primer puesto de las causas de MSC asociada al deporte en España e Italia [Corrado 2011, Suárez 2013], habiéndose asociado la participación en competiciones deportivas, con un desarrollo más precoz de la cardiopatía y a un mayor riesgo de arritmias ventriculares en sujetos afectados por MCA [Ruwald 2015]. Es muy poco frecuente el desarrollo de signos o síntomas de la enfermedad antes o después de los 12 o 60 años de edad respectivamente [Philips 2015].

Asimismo, la MCA engloba un amplio espectro de trastornos sistémicos (como la sarcoidosis o la amiloidosis), inflamatorios, infecciosos, genéticos y de los canales iónicos [Towbin 2019]. Puesto que no todas las personas con la mutación genética desarrollan signos de la enfermedad, es posible encontrar desde pacientes asintomáticos a otros con MSC como primera manifestación clínica. En el medio de estos dos extremos se encuentran muchos pacientes con palpitaciones, síncope o grados variables de insuficiencia cardíaca [Tomé 2004].

Clásicamente se han descrito tres fases de la enfermedad. En la primera fase, denominada temprana o silente, los pacientes habitualmente son asintomáticos, pero presentan riesgo de arritmias ventriculares y MSC. En la segunda fase, denominada eléctrica, suelen aparecer arritmias sintomáticas, normalmente taquicardia ventricular (TV) con morfología de bloqueo de rama izquierda y eje superior, y afectación estructural que no siempre es detectable por técnicas convencionales de imagen. En la tercera fase, denominada avanzada, la progresión de la enfermedad conlleva una disfunción sistólica izquierda, derecha o biV, asociada frecuentemente a arritmias ventriculares [Sanz 2018]. Esta progresión, estructural y eléctrica, incluyendo las arritmias ventriculares malignas, suele presentarse en periodos de tiempo breves conocidos como *fases calientes*, y no de un manera progresiva y continua [Haugaa 2016].

La clasificación actual de la MCA incluye las tres variantes fenotípicas siguientes [Corrado 2017]: 1) variante MCA-VD, es decir, el fenotipo más común caracterizado por la afectación predominante del VD, sin anomalías o con anomalías menores del VI; 2) variante MCA-biV caracterizada por la afectación simultánea del VD y el VI; y 3) variante MCA-VI caracterizada por la afectación predominante del VI, sin anomalías o con anomalías menores del VD.

Los actuales criterios diagnósticos de MCA, revisados recientemente y bautizados como *Padua Criteria* [Corrado 2020], se basan en un enfoque multiparamétrico que abarca anomalías ventriculares funcionales y estructurales, hallazgos de caracterización tisular, alteraciones electrocardiográficas, así como antecedentes familiares y genéticos. Estas anomalías son agrupadas en seis categorías distintas conteniendo criterios que se definen como mayores o menores, cuya combinación establece distintos grados de diagnóstico: definitivo, límite y posible. Para un diagnóstico definitivo, deben cumplirse dos criterios mayores, un criterio mayor y dos menores, o cuatro criterios menores de categorías diagnósticas diferentes. Para un diagnóstico límite debe cumplirse un criterio mayor y un criterio menor, o tres criterios menores de categorías diagnósticas diferentes. Finalmente, para un diagnóstico posible, debe cumplirse un criterio mayor o dos criterios menores de categorías diagnósticas diferentes [Corrado 2020, Sanz 2018].

Mientras que la anterior revisión de estos criterios diagnósticos (*Task Force* de 2010 [Marcus 2010]) estaba únicamente dirigida a la MCA-VD, el *Padua Criteria* actualiza estos criterios y además añade algunos nuevos para el diagnóstico de la MCA-VI. Las novedades aquí consisten fundamentalmente en la introducción de la caracterización tisular por resonancia magnética cardiaca con contraste para la detección de la sustitución miocárdica fibroadiposa de ambos ventrículos. Además, se añaden nuevos criterios ECG, incluyendo algunas anomalías de despolarización, repolarización y arritmias ventriculares específicas para la afectación del VI. No obstante, estos criterios para MCA-VI, en especial los del ECG, son todavía objeto de estudio y deben ser validados más exhaustivamente en estudios clínicos con cohortes más grandes de pacientes [Corrado 2021]. En los pacientes que cumplen los criterios diagnósticos para el fenotipo del VD, la MCA-biV puede diagnosticarse también si aparecen características fenotípicas de la afectación del VI.

Entre las anomalías de la despolarización en el ECG para VD se encuentran las ondas epsilon en las derivaciones precordiales derechas, clasificadas como un criterio ECG menor, al igual que el patrón ECG de una onda S ascendente con QRS retrasado más de 55 ms en las derivaciones precordiales derechas (especialmente si va seguido de ondas T negativas) [Marcus 1982].

Respecto a las anomalías de la despolarización en el ECG para VI, los voltajes QRS bajos en las derivaciones de las extremidades (amplitud QRS pico a pico < 0,5mV) pueden predecir la afectación del VI [Augusto 2020, Segura 2020], considerándose un criterio ECG menor. Esta reducción de voltajes en el QRS ocurre

principalmente por la disminución de la masa miocárdica del VI por sustitución fibroadiposa.

Por otro lado, entre las anomalías de la repolarización del ECG para VI, se describe la inversión de la onda T principalmente en las derivaciones precordiales izquierdas V4-V6, pudiendo predecir afectación MCA-VI [Cipriani 2020, Corrado 2017, Sen 2008], y considerándose un criterio ECG menor. La extensión de la inversión de la onda T desde las derivaciones precordiales derechas (V1-V3) a las derivaciones precordiales izquierdas (V4-V6) puede predecir una dilatación y disfunción más grave del VD, más que la afectación del VI. A modo de ejemplo, la Figura 2.10 presenta las derivaciones precordiales de un paciente diagnosticado por MCA-VD donde se aprecia claramente la inversión de la onda T en las derivaciones V1-V4.

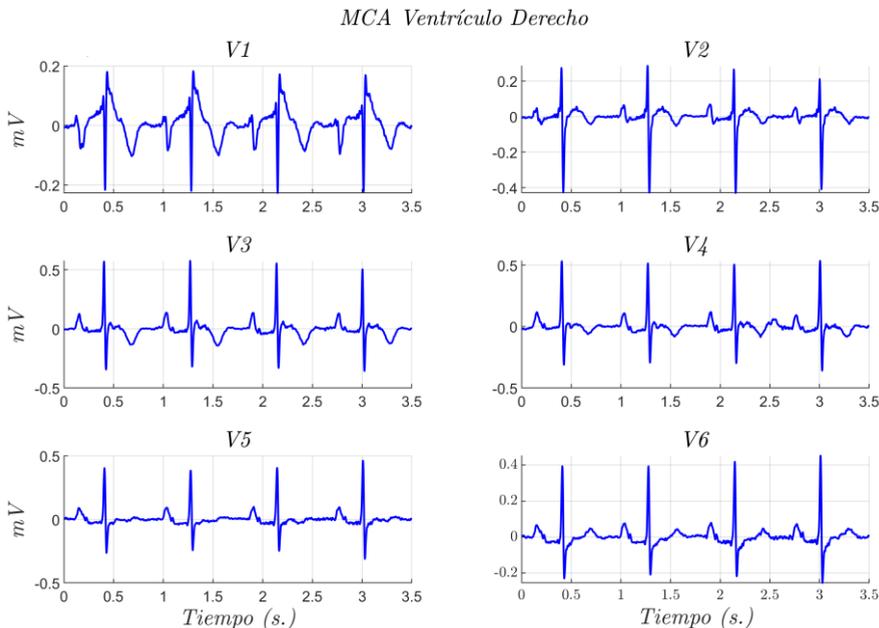


Figura 2.10. Derivaciones precordiales de un paciente afectado por MCA-VD (Hospital Universitari i Politècnic La Fe de València).

En cuanto a las arritmias ventriculares que sirven como criterio diagnóstico encontramos los latidos ventriculares prematuros (LVP), que se evalúan no sólo en términos de número absoluto (más de 500 LVP en 24 h) y complejidad (TV sostenida o no sostenida), sino también con respecto a la morfología del QRS ectópico [Muser 2019]. Por lo tanto, es importante para el diagnóstico registrar la morfología de la arritmia ventricular mediante pruebas de esfuerzo o monitorización Holter de 12 derivaciones ECG durante 24 horas. Para diagnóstico MCA-VD, tanto los LVP, como la TV se consideran criterio mayor si son originados en regiones del VD

distintas del tracto de salida del ventrículo derecho (TSVD), y menor si son originados en el propio TSVD. Por otra parte, las arritmias ventriculares con morfología de bloqueo de rama derecha pueden ser un marcador ECG de afectación del VI, considerándose un criterio de arritmia ventricular menor [Cipriani 2020, Corrado 2017, Sen 2008].

Todo lo indicado respecto los anteriores criterios, sin tener en cuenta las demás técnicas genéticas y de imagen también involucradas, muestra la complejidad del diagnóstico de la enfermedad.

Los tratamientos en el contexto de la MCA incluyen cambios en el estilo de vida, tratamiento farmacológico, ablación con catéter, implante de desfibrilador automático implantable y trasplante cardíaco [Sanz 2018]. Tienen como objetivo principal disminuir la progresión de la enfermedad en pacientes diagnosticados, así como su mortalidad y morbilidad. En este sentido, es sumamente importante el cribado intra-familiar una vez detectado un caso, para detectar posibles nuevos afectados y afectadas.

2.2. Estado del arte

La detección automática de arritmias cardíacas a partir de los datos del ECG, principalmente la FA, se ha estudiado durante las últimas décadas, habiéndose propuesto e implementado una gran variedad de enfoques. En este sentido, la mayoría de metodologías recientes discutidas en la literatura corresponden con algoritmos de clasificación de aprendizaje supervisado, que toman como entrada marcadores o características extraídas previamente del ECG, o más recientemente, la señal ECG directamente. A continuación, se hace una revisión del estado del arte de los distintos marcadores de arritmias existentes, así como de los modelos de clasificación utilizados para detectar anomalías en el ECG. Finalmente, se discute el estado actual de la detección de arritmias en dispositivos móviles.

2.2.1. Marcadores de arritmias basados en la actividad auricular

Los detectores de arritmias, principalmente de FA, basados en el análisis de actividad auricular se fundamentan en la detección de la ausencia de ondas P y/o la presencia de ondas fibrilatorias f que aparecen en el intervalo TQ, señaladas en la Figura 2.11. Para ello, es necesario un análisis de las frecuencias dominantes de la señal ECG, donde aparecen valores elevados debido principalmente a las oscilaciones caóticas de las ondas f , así como a una mayor variabilidad del RR.

Entre los métodos descritos para detectar la FA mediante la actividad auricular encontramos: detección de ausencia de la onda P mediante modelos de mezcla de Gaussianas (GMM) [Ladavich 2015], algoritmos basados en promediado de ondas P [Pürerfellner 2014], uso de la energía y entropía *wavelet* [García 2016, Ródenas 2015], y análisis del número promedio de ondas f [Du 2014].

Es importante destacar que, para permitir un correcto análisis de las ondas fibrilatorias f , es necesario extraer la actividad auricular de la señal ECG cancelando previamente el complejo QRS-T correspondiente a la actividad ventricular. Esta extracción de la actividad auricular requiere de un procesado de señal no lineal, ya que los espectros de la actividad auricular y ventricular se superponen [Rieta 2004]. Para llevar a cabo esta tarea existen diferentes estrategias, como por ejemplo el clásico *Template Matching and Substraction* donde un patrón medio del QRS-T es utilizado para cancelar la actividad ventricular de la señal de ECG [Slocum 1992]. Otro método clásico de extracción de señal auricular es el *Blind Source Separation*, en el que mediante el análisis de componentes principales (PCA) [Raine 2004, Castells 2005a, Castells 2005b], o independientes (ICA) [Rieta 2004], se buscan los diferentes orígenes no correlacionados de la señal auricular y ventricular en el ECG. También se ha utilizado para la extracción de señal auricular, redes neuronales con retardo temporal [Vasquez 2001], con función de base radial (RBF) [Mateo 2013], o métodos basados en *Least Mean Squares* [Ciaccio 2012]. Finalmente, la utilización

de *Echo State Neural Networks* [Petrénas 2012], combinado con el análisis ICA [Mihandoost 2022] parece ser la tendencia actual para la cancelación del complejo QRS-T.

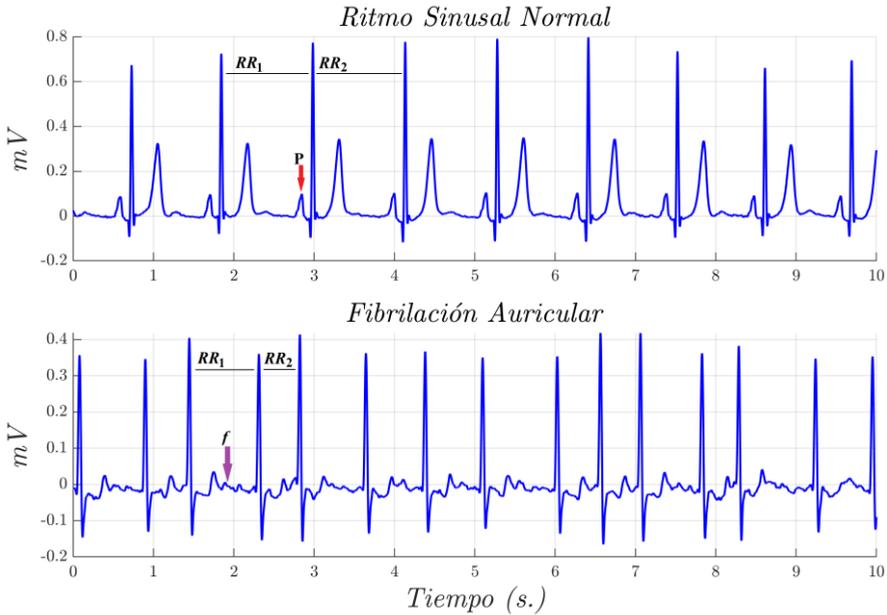


Figura 2.11. Comparación de señales ECG con ritmo sinusal normal frente a fibrilación auricular.

Una vez estimada la onda auricular se procede a realizar el análisis espectral de la señal resultante. Para series temporales de electrogramas auriculares es posible obtener la densidad espectral de potencia, o detectar la frecuencia dominante (FD) que puede servir como marcador de la complejidad de la fibrilación [Bollman 2006]. Las distribuciones de Choi-Williams o de Wigner-Ville también han sido utilizadas en este aspecto para el análisis tiempo-frecuencia [Stridh 2001]. Finalmente, el seguimiento de la evolución de la FD durante la fibrilación puede mejorar su robustez frente al ruido y artefactos utilizando otras técnicas, como los modelos ocultos de Markov (HMM) [Babaeizadeh 2009, Sandberg 2008].

Los detectores de FA basados en el análisis de actividad auricular presentan una elevada precisión si las señales de ECG son poco ruidosas, pero tienden a bajar su rendimiento excesivamente en caso contrario [Colloca 2013a]. Esto es así debido a que la relación señal/ruido de la actividad auricular es muy baja, la detección de ondas *f* o T aisladas para el intervalo TQ-QT es menos eficiente y el rendimiento de los algoritmos de detección de FA basados en la actividad auricular no es óptimo [Asgari 2015, Vásquez 2001]. Asimismo, cabe destacar que los métodos presentados en este apartado requieren de más de una derivación ECG, además de un alto

consumo computacional, con lo que complican su implementación en dispositivos móviles o *wearables*. Es por ello que las características extraídas de las señales ECG durante este trabajo no contemplan aquellas relacionadas con la actividad auricular.

2.2.2. Marcadores de arritmias basados en la actividad ventricular

El análisis de la actividad ventricular en el ECG se basa especialmente en estimar la variabilidad de los intervalos RR entre latidos cardiacos consecutivos, utilizando para ello los sucesivos complejos QRS. Puesto que el QRS presenta la onda de mayor amplitud del ECG, su detección es mucho más robusta frente al ruido [Friesen 1990]. Por lo tanto, este enfoque se considera el más adecuado para la detección automática de arritmias en tiempo real [Carrara 2015].

Existen diversos indicadores para evaluar la Variabilidad de la Frecuencia Cardíaca (HRV, por sus siglas en inglés). Algunos de los métodos descritos para la extracción de marcadores basados en la actividad ventricular incluyen los siguientes análisis o estadísticas respecto a la secuencia de intervalos RR: estadísticas basadas en el tiempo o la frecuencia [Sadr 2018], desviación de la mediana [Linker 2016], primera derivada del histograma [Huang 2011], porcentaje de intervalos que difieren en más de 50 ms (pNN50) [Mietus 2002], detección de irregularidades mediante ventana deslizante en latidos consecutivos [Petrénas 2015a], y relativos a la entropía de Shannon, muestral, normalizada o corregida [Tadi 2019, Zhou 2014, Lim 2016, DeMazumder 2013, Lake 2011, Alcaraz 2010, Anuradha 2008, Cerutti 2008]. Estas técnicas tienen un coste computacional relativamente reducido.

Otros marcadores de arritmias relevantes basados en HRV se derivan del análisis de los gráficos de dispersión de Poincaré o Lorenz [Zhang 2015, Park 2009, Sarkar 2008]. En la Figura 2.12 aparecen algunos ejemplos del gráfico de Poincaré, donde cada punto indica, en un plano 2D, la diferencia de tiempo entre tres latidos consecutivos (RR_n vs RR_{n-1}). En dicha ilustración se muestran 4 tipos de patrones distintos que pueden aparecer: en forma de cometa (panel a), torpedo (panel b), abanico (panel c) y lóbulo lateral doble (LLD, panel d). Los patrones del gráfico de Poincaré en forma de cometa o torpedo se asocian a ritmo sinusal normal, el patrón en abanico a la FA, mientras que patrones de tipo LLD se han descrito en ritmos sinusales, pero con contracciones auriculares y ventriculares prematuras. No obstante, existen más tipos de patrones descritos también en [Esperer 2008]. Las señales utilizadas en este ejemplo contenían 24 horas de monitorización adquiridas mediante un dispositivo *Holter*, lo que propicia el elevado número de puntos en las gráficas, situación que no es posible alcanzar con registros cortos de ECG.

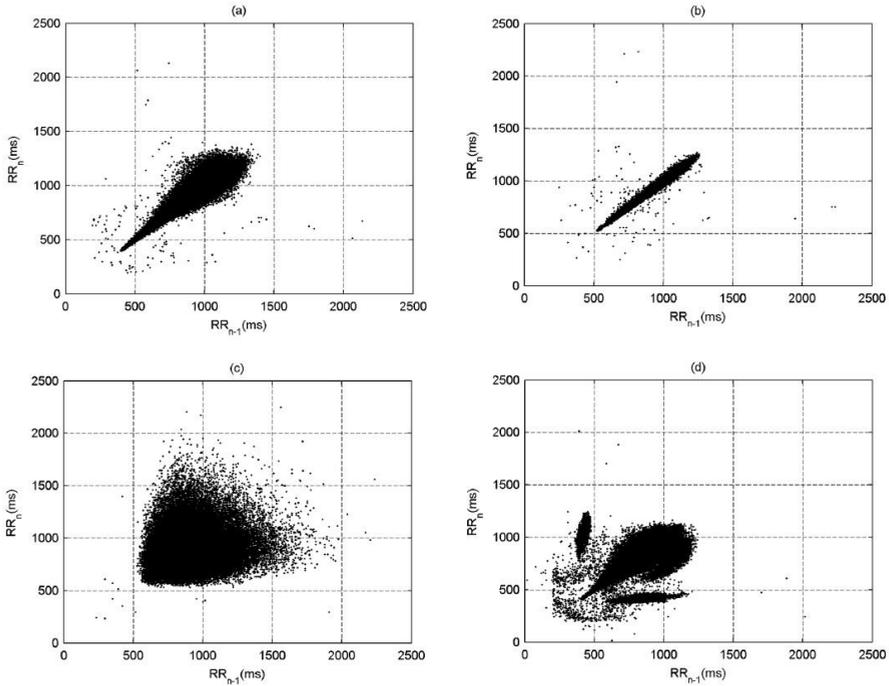


Figura 2.12. Gráficos de Poincaré [Zhang 2015].

Ejemplos de 4 tipos de patrones distintos: a) cometa, b) torpedo, c) abanico, d) lóbulo lateral doble. Todos ellos, generados con registros ECG de dispositivo *Holter* de 24 horas de duración.

También es común el uso de marcadores que caracterizan la duración de los distintos segmentos del ECG. Aquí, un marcador destacado es el QT corregido (QTc), parámetro utilizado para ajustar la duración del intervalo QT en función de la frecuencia cardiaca, siendo un indicador relevante para evaluar el riesgo de arritmias. Por ejemplo, una prolongación del QT predispone a muerte súbita por arritmias ventriculares malignas como *torsade de pointes* [Medeiros 2007]. Hay varias fórmulas utilizadas para calcular el QTc, donde QT y RR indican la duración de los intervalos en ms, siendo algunas de las fórmulas más comunes las de Bazzet ($QTcB$), Fridericia ($QTcF$) y Framingham ($QTcG$) [Bazzet 1920, Fridericia 2003, Sagie 1992], detalladas a continuación.

$$QTcB = QT/\sqrt{RR} \tag{2.1}$$

$$QTcF = QT/\sqrt[3]{RR} \tag{2.2}$$

$$QTcG = QT + 0.154(1 - RR) \tag{2.3}$$

La elevación del segmento ST es otro marcador indicativo de diferentes patologías como la isquemia transmural o el infarto de miocardio entre otros [Bliet 2018, Desphande 2014]. Los métodos utilizados para caracterizar esta elevación van desde la extracción de características morfológicas del ECG, transformadas de Fourier, Karhunen-Loève y *wavelets*, análisis de componentes principales, e incluso el análisis del vectocardiograma [Ansari 2017].

Otros métodos se centran en la detección de latidos ectópicos, o extrasístoles, mediante distintas técnicas revisadas en [Nabil 2015]. Ahí, se contemplan, de menor a mayor complejidad computacional, aproximaciones basadas en la duración y morfología de complejos semejantes en amplitud al QRS, transformaciones morfológicas de la señal como la dilatación y la erosión, transformadas *wavelets*, métodos de descomposición modal empírica o redes neuronales.

También cabe señalar los distintos marcadores utilizados por los dispositivos desfibriladores para la detección del pulso ante una parada cardíaca súbita o fibrilación ventricular. Esta detección es esencial para reconocer el evento, así como el retorno a un ritmo cardíaco adecuado después del choque eléctrico suministrado, crucial para la supervivencia del paciente. Por ejemplo, [Elola 2018] utilizó un clasificador *random forest* que tomaba como entrada nueve características procedentes del análisis del tiempo, frecuencias, pendientes y regularidad del ECG. Este método logro alcanzar una sensibilidad media del 88,4 % en segmentos de ECG de 4 segundos de duración, pudiendo integrarse en cualquier desfibrilador externo automatizado comercial. A este respecto, [Barker 2022] revisó recientemente los métodos automáticos de detección de riesgo de muerte súbita para dispositivos cardioversores implantables, donde la mayoría de ellos tenían el mismo enfoque de extracción de características temporales y frecuenciales del ECG junto con un clasificador basado en aprendizaje automático.

Cabe señalar que los detectores de arritmias que combinan tanto la actividad auricular como la respuesta ventricular pueden mejorar el rendimiento de clasificación al combinar datos independientes de cada parte del ciclo cardíaco. Por ejemplo, en [Petrénas 2015b] se aplicó un método de clasificación por lógica difusa combinando la irregularidad del intervalo RR, la ausencia de onda P, la presencia de ondas *f* y el nivel de ruido. También existen enfoques multivariantes que combinan varias de las características individuales anteriores y que proporcionan un rendimiento adecuado en la detección de FA [Li 2016, Oster 2015, Colloca 2013b].

Además de los métodos mencionados anteriormente, entre otras técnicas avanzadas también utilizadas para la extracción de características del ECG encontramos la densidad espectral de potencia, la matriz de co-ocurrencia, las transformadas de Fourier de tiempo reducido, o las estadísticas de orden superior [Marinho 2019, Nascimento 2019, Hussein 2018]. Otros métodos no lineales de extracción de características son los gráficos de recurrencia, la dimensión fractal, el método de Burg, el exponente de Lyapunov, el análisis de fluctuación de tendencia y el análisis de la dimensión de correlación [Lim 2016]. Los métodos de búsqueda

de similitud entre dos formas de onda, aplicados a la transformada *wavelet*, también son eficaces para la extracción de características del ECG [Kora 2016]. Algunas de estas técnicas mejoran el rendimiento de los modelos al utilizarse conjuntamente; por ejemplo, se ha observado que el uso de la matriz de co-ocurrencia junto con estadísticas de orden superior funcionan mejor cuando se aplican simultáneamente [Nascimento 2019]. Sin embargo, la extracción de estas características complejas tiene un coste computacional algo más elevado que las mencionadas inicialmente.

El bajo coste computacional y la robustez frente al ruido son requisitos clave si tenemos en cuenta futuras implementaciones de la extracción de marcadores en dispositivos de bajo consumo energético para la detección de arritmias. En este sentido, autores como [Kora 2017] y [Annarapu 2016] propusieron y compararon diferentes versiones de métodos basados en la transformada compleja de Hadamard para la extracción de características ECG de latidos individuales. Estos métodos presentaron un bajo consumo computacional y de memoria, y la generación de características relevantes y no redundantes; todo ello superando a otros enfoques en términos de sensibilidad, especificidad y precisión de detección global de FA al ser utilizados por distintos modelos de clasificación.

Como ha quedado demostrado, la investigación de marcadores de arritmias es muy amplia, y en cada patología se suele hacer uso de aquellos parámetros más indicados para su detección.

Nuestra metodología aborda la obtención de marcadores del ECG principalmente desde el análisis de la actividad ventricular mediante la extracción de características sencillas. Las dos razones principales para esto son: 1) la señal auricular normalmente está contaminada por ruido o artefactos en dispositivos de una única derivación, y 2) el coste computacional de detectar las ondas ventriculares es considerablemente menor que el necesario para la extracción y el análisis de la actividad auricular, o la obtención de características ventriculares complejas. Esto es especialmente relevante si se pretende implementar y evaluar metodologías susceptibles de ser utilizadas en dispositivos con pocos recursos computacionales, como dispositivos móviles o *wearables*.

2.2.3. Aprendizaje automático para el diagnóstico de arritmias

Los primeros enfoques para la detección automática de anomalías en el ECG se basaron especialmente en la caracterización de ritmos a partir de la detección del QRS, el análisis tiempo-frecuencia y sus características, o en las transformadas de señal de Fourier y *wavelets* [Alexakis 2003, Chazal 2004, Christov 2006, Mahmoud 2006, Kostka 2007, Minami 1999, Martínez 2004, Mahmoodabadi 2005, Alcaraz 2006, Yang 2013, Aqil 2015]. Sin embargo, actualmente estas técnicas se combinan con otras de aprendizaje automático (AA), alcanzando así mejores resultados y

siendo el estándar actual en la detección de arritmias [Zhang 2023, Pujadas 2022, Rizwan 2021].

A este respecto, el AA, se centra en el diseño y desarrollo de modelos matemáticos y estadísticos que ayudan a la toma de decisiones de forma automatizada basándose en la inferencia adquirida a partir de los datos proporcionados para su entrenamiento [Alpaydin 2016, Murphy 2012]. Dentro del AA encontramos las técnicas de reconocimiento de patrones, que tienen como objetivo clasificar o agrupar muestras desconocidas en diferentes categorías mediante un modelo entrenado previamente con un conjunto de muestras iniciales [Bishop 2006].

Convencionalmente, el AA se divide en aprendizaje supervisado y no supervisado. El aprendizaje supervisado se ocupa de problemas en los que se conoce la salida, y se divide en algoritmos de regresión cuando se busca predecir valores numéricos continuos, o algoritmos de clasificación cuando el objetivo es identificar variables categóricas. Por ejemplo, para determinar si un registro ECG es normal o presenta un episodio de arritmia, donde se conoce la clase o etiqueta (normal o arritmia), se utilizaría un modelo de clasificación. Por el contrario, en el caso del aprendizaje no supervisado, se desconocen las etiquetas o clases, siendo las técnicas de agrupación o *clustering* el ejemplo más clásico y utilizado.

Las soluciones basadas en el AA para detección de anomalías en el ECG, disponibles en la literatura, realizan principalmente una clasificación para distinguir el escenario del ritmo sinusal normal del de distintas arritmias. Se puede encontrar una extensa revisión reciente de los algoritmos más utilizados en [Denysyuk 2023], entre los que se encuentran los siguientes: la clasificación por *Naïve Bayes* (NB), el análisis de discriminación lineal, los K vecinos más cercanos (KNN), los modelos ocultos de Markov (HMM) [Sandberg 2008], las redes neuronales artificiales (RNA), la regresión logística (RL), las *Support Vector Machines* (SVM) [Yin 2019], los árboles de decisión [Zabih 2017], los algoritmos de similitud secuencial, los algoritmos de teoría de grafos [Luz 2013] o la clasificación basada en reglas.

Asimismo, los enfoques basados en el AA pueden agruparse en dos categorías principales: 1) los que utilizan características de la señal ECG como entrada, y 2) aquellos que utilizan los datos crudos (*i.e.* la propia señal ECG) como entrada.

En el primer enfoque, se utilizan como entrada las características extraídas de los registros ECG. Los algoritmos clásicos como NB, SVM, KNN, los árboles de decisión, o las RNA de tipo *Feed Forward* (FFNN) siguen este enfoque. El uso de estos algoritmos implica el paso computacional adicional de la extracción de características. No obstante, su coste computacional durante la etapa de clasificación o inferencia suele ser reducido, sobre todo si se utiliza un conjunto pequeño y adecuado de características, habitualmente seleccionado en etapas previas. Un ejemplo de esta estrategia lo encontramos en [Hagan 2021], donde se probaron SVM, FFNN, y conjuntos de árboles de decisión para la clasificación de cinco arritmias diferentes, siendo los árboles los que mejor relación entre precisión y coste

computacional mostraron. Otro ejemplo lo encontramos en [Zheng 2020a], donde un clasificador de tipo *Gradient Boosting Tree* logró el mejor rendimiento de clasificación de registros ECG entre los cuatro ritmos cardíacos siguientes: ritmo sinusal, FA, bradicardia sinusal y taquicardias auriculares o ventriculares.

El segundo enfoque no requiere de la extracción de características ya que se utilizan únicamente los datos brutos como entrada. Esto es así principalmente gracias a las redes neuronales convolucionales (CNN), cuya aproximación se basa no solo en aprender de las características de entrada, sino generar estas mismas características a base de convoluciones en los datos crudos. Estas nuevas versiones avanzadas de las RNA, se conocen como aprendizaje profundo o *Deep Learning* por su elevado número de capas de neuronas interconectadas (unidades computacionales). Actualmente son muy populares por su eficacia en distintos ámbitos de clasificación, principalmente de imágenes. Sin embargo, estos enfoques tienen un coste computacional muy elevado, sobre todo durante la etapa de entrenamiento, debido principalmente al gran número de nodos y sus correspondientes parámetros a estimar. Este elevado coste computacional dificulta su implementación en dispositivos portátiles. Para sortear este inconveniente, se han planteado soluciones basadas en computación en la nube, donde las señales se adquieren en dispositivos portátiles, pero se procesan y evalúan en remoto [Hartmann 2019]. En otros casos, como por ejemplo dentro de instalaciones hospitalarias, es posible implementar soluciones basadas en *edge computing*, procesando los datos cerca de su origen, reduciendo así el tiempo de transferencia de datos y el coste computacional para los dispositivos portátiles [Sodhro 2019]. No obstante, estas soluciones no resolverían el problema para dispositivos no conectados a la red de datos, o en lugares con dificultades de conexión.

Podemos encontrar algunos ejemplos recientes que han utilizado este último enfoque en [Krusteva 2021, Rahimeh 2021, Xiong 2018], donde se emplearon CNN para detectar FA en una única derivación. Otros trabajos tenían como objetivo detectar hasta nueve afecciones cardíacas en bases de datos de ECG extensas utilizando CNN [Dai 2021, Hua 2021, Jo 2021, Yang 2021, Zhang 2021]. Otros autores publicaron resultados sobre la detección de 24 a 27 anomalías cardíacas en bases de datos que contenían un máximo de 66.361 muestras, utilizando 12 derivaciones y diferentes *frameworks* basados en CNN [Giovanni 2021, Zhu 2021, Zhao 2022]. Recientemente, durante el *PhysioNet/Computing in Cardiology Challenge 2021* [Pérez 2020, Reyna 2021], se hizo pública una base de datos con 88.253 registros de ECG y 26 tipos diferentes de arritmias. Hasta la fecha, las aproximaciones que mejores rendimientos de clasificación han obtenido están basadas en aprendizaje profundo [Antoni 2022, Nejedly 2022, Xu 2022].

Puesto que los enfoques basados en AA necesitan de conjuntos grandes de datos para su entrenamiento, sobre todo los basados en aprendizaje profundo, las enfermedades raras como la MCA suelen estar fuera del alcance de estos trabajos en la literatura.

Finalmente, cabe destacar que la detección del complejo QRS es un paso preliminar obligatorio en casi todas las soluciones basadas en AA para el diagnóstico de arritmias en el ECG. En este aspecto, encontramos como inconveniente que los registros ECG suelen contener distintos tipos de ruido acoplado a la señal: interferencia de la red eléctrica, electromiográfica, artefactos por movimientos bruscos, o desviación de la línea base debido a la respiración. La presencia de ruido afecta a la detección del complejo QRS y, en última instancia, a la extracción de características y al rendimiento del modelo de clasificación. En consecuencia, es necesario remarcar la dependencia de los distintos enfoques basados en AA de la calidad de las señales que reciben como entrada, así como de la resistencia al ruido de los algoritmos de detección del complejo QRS [Friesen 1990], y de extracción de características. Algunos trabajos recientes, como [Lastre 2019], demostró que un filtrado adecuado, como el adaptativo de tipo UFIR (*unbiased FIR*), mejoran la calidad de las características extraídas de la señal ECG, resultando más fiables y menos propensas a grandes desviaciones de los valores medios.

2.2.4. Detección de arritmias en dispositivos móviles

Los dispositivos portátiles están ganando un gran interés para la detección temprana de enfermedades cardíacas, tanto en la investigación como en entornos clínicos, debido a su potencial para detectar y posibilitar el seguimiento masivo de distintas afecciones en una amplia variedad de escenarios [Hermans 2022, Duncker 2021, Kumari 2017]. Un ejemplo de escenario actual es el cribado mediante ECG, el cual ya se sugiere como una estrategia para aumentar las tasas de detección de FA e iniciar los tratamientos en una fase más temprana en individuos de alto riesgo [Fan 2018].

Este cambio de paradigma hacia la monitorización ambulatoria y automonitorización a través de dispositivos portátiles ha introducido requisitos adicionales, como la detección en tiempo real, la capacidad de hacer frente a señales ruidosas, de baja calidad y con artefactos, además de duras restricciones en la complejidad computacional y el consumo de energía [Rizwan 2021]. Al mismo tiempo, los investigadores están aprovechando esta nueva y extensa fuente de datos para obtener información sobre lo que les ocurre a los pacientes fuera del ámbito hospitalario [Harmon 2023]

Entre los dispositivos portátiles o *wearables* para el diagnóstico de arritmias, encontramos el *Kardia Mobile* de AliveCor, el *ECG Check* de Cardiac Designs, el *Nuubo System* de la empresa española Nuubo, o el *Apple Watch ECG* [Bumgarner 2018]. Todos estos dispositivos cuentan con la aprobación de organismos regulatorios como la *Food and Drug Administration* (FDA) para el registro y el diagnóstico de arritmias, mayoritariamente FA [Harmon 2023, Garabelli 2017, Isakadze 2020]. Algunos de estos productos se basan en relojes inteligentes, mientras que otros se basan en un dispositivo externo con diferentes electrodos para adquirir

la señal ECG, mediante un número reducido o único de derivaciones, normalmente equivalente a la derivación I.

Dado el creciente uso de estos aparatos, con venta directa al consumidor, se plantean importantes cuestiones sobre su utilización en la práctica clínica, existiendo actualmente un amplio debate sobre sus ventajas e inconvenientes.

Respecto a las ventajas principales destaca la accesibilidad generalizada a la monitorización ECG que permite introducir esta tecnología en dispositivos relativamente baratos o compatibles con teléfonos móviles [Lawin 2022, Garabelli 2017]. Aunque estos dispositivos no sean tan sofisticados y exactos como los ECG clínicos, presentan la ventaja de poder registrar largos periodos de señal [Pujada 2022]. Esto posibilita la detección temprana de arritmias junto con un diagnóstico más precoz, pudiendo ayudar en la promoción de la salud y estrategias preventivas. También facilitaría el tratamiento y seguimiento de enfermedades, permitiendo asumir a los pacientes un papel más activo en el cuidado de su propia salud [Isakadze 2020]. Otra ventaja es que pueden descentralizar la recopilación de señales ECG, mitigando así las limitaciones de los ensayos clínicos convencionales [Harmon 2023]. Asimismo, la detección en tiempo real de patrones arrítmicos malignos permitiría dar la alarma inicial a los servicios de emergencia o a las personas del entorno [Rizwan 2021]. Todos los escenarios anteriores son especialmente interesantes en zonas rurales aisladas o en situaciones de pandemia o catástrofes [Spaccarotella 2020].

No obstante, la implementación de la tecnología móvil en la prevención y tratamiento de enfermedades cardiovasculares todavía enfrenta inconvenientes y retos relevantes. Se requiere más investigación y colaboración entre los organismos reguladores y las empresas tecnológicas para resolver la falta de información sobre resultados, falsos positivos y privacidad de datos [Isakadze 2020]. Además, la eficacia de estos dispositivos aún debe ser establecida [Rizwan 2021]. Varios autores sugieren que, aunque los dispositivos de ECG portátiles son útiles para detectar la FA, se necesitan más estudios para demostrar y mejorar el diagnóstico, así como para expandir su uso a otras indicaciones y evaluar sus efectos positivos en la atención clínica [Lawin 2022, Al-Alusi 2019]. En general, estos avances contribuirían a mejorar su aceptación en la práctica clínica. Mientras tanto, esta tecnología debe utilizarse con precaución y criterio, estando especialmente indicada en pacientes de alto riesgo que no tengan acceso a una valoración médica rápida, entrenados en la técnica y con posibilidad de contacto telemático con un centro sanitario.

En último lugar, es necesario mencionar que existe la necesidad de una investigación y desarrollo más profundo para la detección en tiempo real de arritmias. La mayoría de estudios se basan en el procesamiento *a posteriori* de las señales ECG ya guardadas en bases de datos, desarrollando modelos de AA donde se evalúa su rendimiento de clasificación, pero no su coste computacional o posibilidad de implementación en dispositivos portátiles [Parvaneh 2019].

Capítulo 3

Materiales y métodos generales

En esta sección se presentan los materiales y métodos comunes utilizados en diferentes capítulos de esta tesis. Se introducirá brevemente los fundamentos teóricos de los distintos tipos de modelos de clasificación que se utilizaron durante la experimentación. En particular, se describen cuatro tipos diferentes de clasificadores: redes neuronales artificiales (RNA) de tipo *Feed Forward* (FFNN), Naïve Bayes (NB), *Support Vector Machines* (SVM) y regresión logística (RL). Asimismo, se describen los conceptos de sensibilidad y especificidad utilizados. Los materiales y métodos específicos se describirán en cada capítulo.

3.1. Aprendizaje automático

El aprendizaje automático (AA) es un conjunto de técnicas que permiten a los sistemas informáticos aprender a partir de los datos suministrados, produciendo modelos o estructuras de datos que se adaptan de forma progresiva a dichos datos. Por otro lado, el reconocimiento de patrones se refiere al proceso de identificar características relevantes en un conjunto de muestras para clasificarlas en diferentes categorías [Bishop 2006]. Para lograr este objetivo, se utilizan modelos matemáticos o estructuras de datos complejas que permiten etiquetar una nueva muestra de datos en la clase correspondiente. Por lo tanto, para generar modelos de reconocimiento de patrones se requieren los procesos matemáticos y computacionales del AA.

El proceso de generación de un modelo para el reconocimiento de patrones basado en aprendizaje automático se divide en dos fases principales: entrenamiento y reconocimiento. Durante la fase de entrenamiento se utiliza un conjunto de datos para construir el modelo, llamado conjunto de entrenamiento. Es en esta fase donde se ajustan los parámetros de un modelo adaptativo para obtener la mejor generalización posible, y de ese modo resolver nuevos casos durante la fase de reconocimiento. Una vez el modelo está listo, es posible utilizarlo para identificar y clasificar nuevas observaciones. La Figura 3.1 muestra un esquema básico del

proceso de generación de un modelo para el reconocimiento de patrones, utilizando un enfoque de aprendizaje automático supervisado.

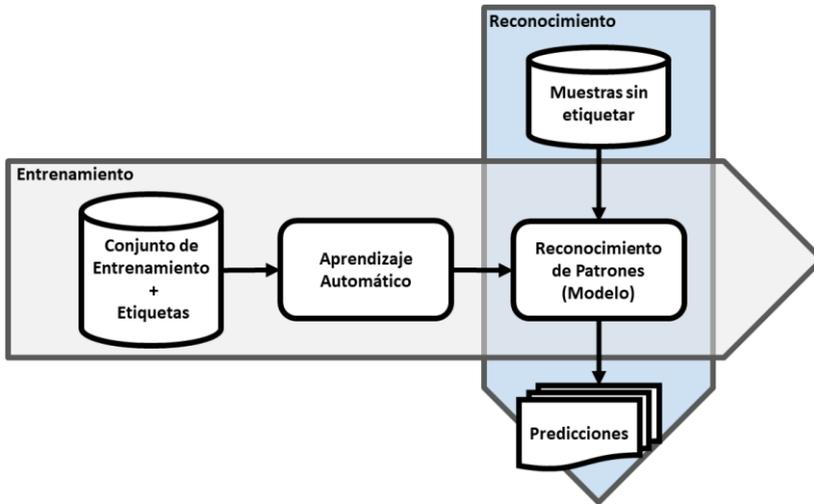


Figura 3.1. Proceso de generación de un modelo para el reconocimiento de patrones.

El objetivo del AA es desarrollar un modelo o regla de decisión M que pueda asignar una clase del conjunto C a un vector de características del conjunto X^D , donde D representa la dimensionalidad o número de características que aparecen en cada una de las muestras [Tortajada 2012]. En resumen, un modelo es una asignación que se presenta en la forma

$$M: X^D \rightarrow C \quad (3.1)$$

Un conjunto de entrenamiento S viene determinado por los vectores de características $x \in X^D$ y sus respectivas clases $c \in C$, produciendo muestras $s_i = \{x_i, c_i\}$. Dicho conjunto de entrenamiento puede expresarse de la siguiente forma

$$S = \{(x_1, c_1), (x_2, c_2), \dots, (x_N, c_N)\} \in (X^D \times C) \quad (3.2)$$

La probabilidad $p(x_i, c_i)$ asocia muestras y clases, y viene determinada por la ecuación 3.3, donde el valor $p(c)$ indica la probabilidad *a priori* de la clase, mientras que el valor $p(x|c)$ es conocido como la probabilidad condicional de la clase, o simplemente la probabilidad condicional.

$$p(x_i, c_i) = p(x_i)p(c_i|x_i) = p(c_i)p(x_i|c_i) \quad (3.3)$$

Generalmente, el modelo se puede definir mediante la siguiente función parametrizada, que intenta aproximar el valor de c , y donde α representa el conjunto de parámetros a estimar del modelo

$$\hat{c} = f_M(x, \alpha) \quad (3.4)$$

Dada la anterior expresión, donde \hat{c} es la estimación de c , es posible cuantificar los errores cometidos en dicha estimación mediante una función de pérdida, comúnmente conocida como *loss function*, y definida en un problema de clasificación del siguiente modo

$$L(c, \hat{c}) = \begin{cases} 0 & \text{si } \hat{c} = c \\ 1 & \text{si } \hat{c} \neq c \end{cases} \quad (3.5)$$

De esta manera, es factible calcular el riesgo condicional del modelo M , expresado como

$$R(c|x) = \sum_{c \in \mathcal{C}} L(c, f_M(x, \alpha)) p(c|x) \quad (3.6)$$

Se define como regla de decisión óptima aquella que consigue una mínima probabilidad de error. Si se conocen las probabilidades *a priori* y condicionales de la clase, la regla de decisión óptima es la regla de decisión Bayesiana expresada como

$$\hat{c} = \arg \min_{c \in \mathcal{C}} R(c|x) \quad (3.7)$$

Por otro lado, cuando se asume la función de pérdida L , entonces el riesgo condicional es el error de probabilidad medio, el cual se expresa como

$$R(\hat{c}|x) = 1 - p(\hat{c}|x) \quad (3.8)$$

Sin embargo, es frecuente que las distribuciones de probabilidad sean desconocidas y se deban aproximar a través del conjunto de entrenamiento \mathcal{S} , el cual se presume que ha sido generado siguiendo $p(x, c)$. En un proceso de AA se asume que tanto las muestras conocidas como las desconocidas son generadas por el mismo proceso estocástico, lo que implica que los datos muestreados de forma aleatoria pertenecen a la misma distribución de probabilidad. En consecuencia, basándose en el conjunto de muestras conocidas \mathcal{S} , durante el aprendizaje automático se trata de encontrar una función o modelo M cuyo riesgo sea lo más cercano posible a $R(\hat{c})$.

Puesto que el modelo $f_M(x, \alpha)$ es estimado con el conjunto de entrenamiento ya etiquetado S , la estimación del rendimiento del modelo se cuantifica utilizando el riesgo empírico R_{emp} definido como

$$R_{emp}(M_\alpha) = \frac{1}{N} \sum_{i=1}^N L(c_i, f_{M_\alpha}(x_i, \alpha)) \quad (3.9)$$

Por otro lado, cuando se asume la función de pérdida L , R_{emp} viene determinado por la siguiente expresión

$$R_{emp}(M_\alpha) = \frac{1}{N} \sum_{i=1}^N \delta(c_i, \hat{c}_i) \quad (3.10)$$

donde δ hace referencia a la regla delta de Kronecker definida como

$$\delta(i, j) = \begin{cases} 0 & \text{si } i = j \\ 1 & \text{si } i \neq j \end{cases} \quad (3.11)$$

En resumen, durante el proceso de entrenamiento de un modelo para el reconocimiento de patrones, se recurre a algoritmos de AA que ajustan los parámetros del modelo en función de la minimización del riesgo condicional. Si bien la selección de los algoritmos dependerá del clasificador a entrenar, sus fundamentos teóricos son comunes a todos ellos.

3.1.1. Redes Neuronales Artificiales

Las redes neuronales artificiales (RNA) son un tipo de modelo para el reconocimiento de patrones inspirado en la estructura y funcionamiento del cerebro humano. Estas redes consisten en un conjunto de unidades de procesamiento simples, denominadas neuronas o nodos, que trabajan de manera conjunta para llevar a cabo tareas complejas de procesamiento de información. Las neuronas están interconectadas por medio de conexiones ponderadas que permiten la transmisión de señales entre ellas.

El modelo básico de una neurona artificial está basado en el perceptrón inicialmente descrito en [Rosenblatt 1958]. Este modelo matemático representa la forma más simple de una neurona, que recibe entradas (señales) y las combina mediante un conjunto de pesos para producir una salida. El proceso de aprendizaje de las redes neuronales se basa en la adaptación de estos pesos para que la red pueda ajustarse a los datos de entrenamiento y mejorar su capacidad para realizar predicciones precisas [Gondra 2008].

Un perceptrón produce una combinación lineal (y) de los valores de sus n entradas (x_i), cada una con su respectivo peso (w_i), sumando un valor de sesgo o *bias* (b). A continuación, se aplica una función de activación (f) a la combinación lineal de pesos y entradas, resultando en la salida del perceptrón expresada mediante la siguiente ecuación:

$$y = f\left(\sum_{i=1}^n x_i w_i + b\right) \quad (3.12)$$

En la Figura 3.2 se ilustra un diagrama básico que describe el funcionamiento de un perceptrón de estas características.

Si las muestras que pertenecen a dos clases diferentes son linealmente separables, un solo perceptrón puede clasificarlas con precisión trazando una frontera mediante un hiperplano que permita distinguir cada una de las dos clases [Hu 2001].

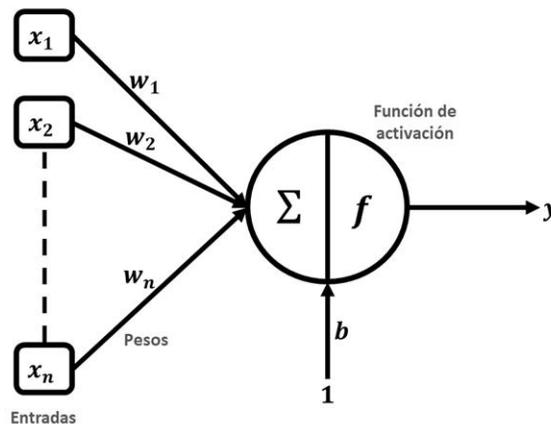


Figura 3.2. Esquema básico de un perceptrón con n entradas.

Aunque una sola unidad tiene una capacidad de computación limitada, un conjunto de neuronas interconectadas posee una gran capacidad de adaptación y la habilidad de aprender funciones no lineales, permitiendo modelar relaciones complejas entre las entradas y salidas durante el proceso de entrenamiento. La construcción de estas funciones más generales es posible mediante redes con sucesivas capas de perceptrones conectados entre sí, dando lugar al perceptrón multicapa comúnmente conocido como *Feed Forward Neural Network* (FFNN). Un FFNN consta de una capa de entrada, una o dos capas ocultas y una capa de salida, que corresponde a una unidad por cada clase en un problema de clasificación. Si todas las unidades de cada capa reciben como entrada la salida de todas las unidades de la capa anterior, y cada salida se conecta con todas las unidades de la siguiente

capa, se considera que está totalmente conectado. La Figura 3.3 muestra un ejemplo de RNA de tipo FFNN, compuesta de una capa de entrada con tres unidades, dos capas ocultas de cinco unidades cada una, y una capa de salida con una única unidad que constituye la respuesta del modelo.

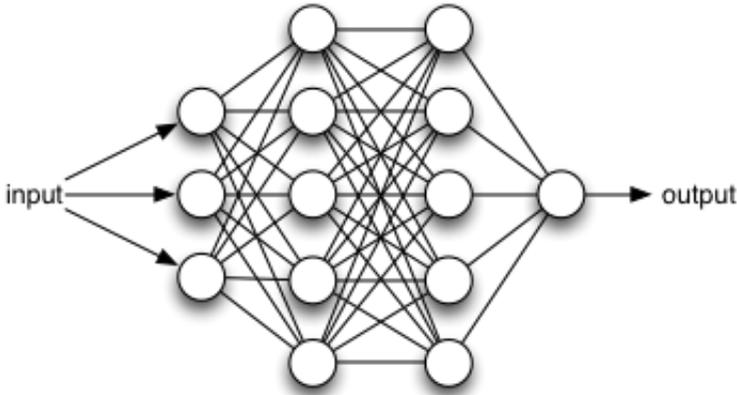


Figura 3.3. Ejemplo de arquitectura de una red neuronal artificial de tipo *Feed Forward*.

Durante la experimentación de esta tesis se utilizaron FFNN totalmente conectadas con una capa oculta. Además, se exploraron modelos con diferentes números de unidades en la capa oculta, detallando los valores exactos utilizados en la metodología del capítulo correspondiente.

Respecto a las funciones de activación en las unidades de la capa oculta, se exploraron modelos con las funciones *tansig* y la función ReLU (*Rectified Linear Unit*), definidas según las siguientes expresiones:

$$tansig(x) = \frac{2}{(1 + e^{-2x}) - 1} \quad (3.13)$$

$$ReLU(x) = \max(0, x) \quad (3.14)$$

Las derivadas de estas funciones son utilizadas durante el proceso de entrenamiento, y están definidas según las siguientes ecuaciones:

$$tansig'(x) = 1 - tansig(x)^2 \quad (3.15)$$

$$\text{ReLU}'(x) = \begin{cases} 1 & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases} \quad (3.16)$$

La función *tansig* (tangente sigmoide) transforma la entrada de la neurona en un valor entre -1 y 1, con una zona de transición suave alrededor de cero, ayudando a normalizar los datos de entrada en un rango acotado, facilitando así el procesamiento de los mismos por la red neuronal. Una de las ventajas de esta función de activación es que permite aprender valores de entrada negativos, posibilitando modelar funciones y relaciones no lineales más complejas. No obstante, su coste computacional es relativamente elevado si la red contiene muchos nodos.

Por otro lado, existen varias ventajas de utilizar la función ReLU. En primer lugar, esta función ayuda a reducir la carga computacional ya que su cálculo y el de su derivada no requieren de la realización de exponenciales o divisiones, permitiendo así un procesamiento más eficiente. Además, gracias a la forma simple de la derivada de la función para valores positivos, se consigue una convergencia más rápida durante el entrenamiento, teniendo implicaciones positivas para la reestimación de pesos [Zeiler 2013]. Sin embargo, esta misma circunstancia también es su debilidad, ya que cualquier entrada negativa a la función ReLU tiene como salida un valor nulo, lo que impide aprender correctamente valores negativos.

Las diferencias entre las funciones de activación descritas pueden apreciarse en la Figura 3.4.

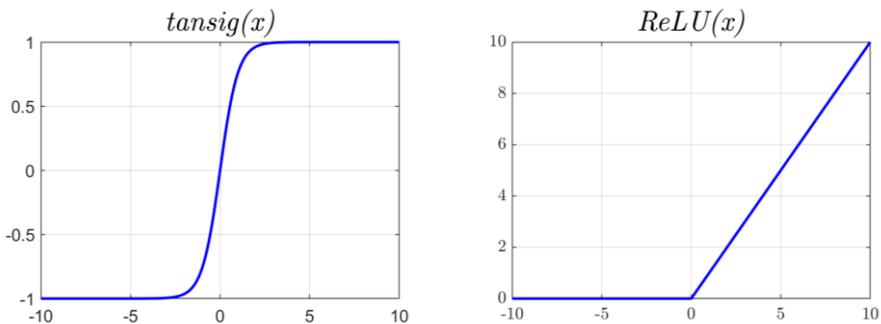


Figura 3.4. Funciones de activación utilizadas para las unidades de la capa oculta.

Para las unidades de las capas de salida, exploramos las funciones de activación *softmax* para el caso de modelos que tenían como salida más de una clase, y *mapminmax* para los modelos FFNN de respuesta binaria.

La expresión matemática de la función de activación *softmax* aparece en la ecuación 3.17, donde x es el vector de entrada, k es el número de clases y i es el índice de la clase actual:

$$\text{softmax}(x)_i = \frac{e^{x_i}}{\sum_{j=1}^k e^{x_j}} \quad (3.17)$$

La función *softmax* normaliza los valores de entrada en una distribución de probabilidades que suman 1, permitiendo su uso en problemas de clasificación multiclase.

Por otro lado, la expresión matemática de la función de activación *mapminmax* aparece en la ecuación 3.18, donde x es el vector de entrada, y_{max} y y_{min} son los valores máximo y mínimo del rango deseado de salida, y i es el índice de la clase actual:

$$\text{mapminmax}(x)_i = (y_{max} - y_{min}) \frac{(x_i - x_{min})}{(x_{max} - x_{min})} + y_{min} \quad (3.18)$$

Aunque *mapminmax* y *softmax* son muy diferentes en su estructura y aplicación, ambas son utilizadas comúnmente para normalizar valores de entrada y producir valores de salida escalados que se ajustan a un rango específico, evitando así que algunos valores dominen sobre otros. En general, ambas funciones de activación son herramientas útiles para mejorar el rendimiento y la precisión de una red neuronal.

Por último, el *learning rate* y el *momentum* son dos hiperparámetros fundamentales en el entrenamiento de redes neuronales de tipo FFNN. El *learning rate* determina cuánto se deben actualizar los pesos de la red durante el proceso de entrenamiento, mientras que el *momentum* agrega una fracción de la actualización anterior para ayudar a acelerar el entrenamiento y evitar que la red quede atrapada en mínimos locales. Una elección adecuada de estos hiperparámetros puede tener un impacto significativo en la velocidad y precisión del entrenamiento, ya que un *learning rate* demasiado alto puede hacer que la red no converja, mientras que un *momentum* inadecuado puede provocar un sobreajuste o subajuste de la red. Por lo tanto, es importante elegir cuidadosamente los valores de estos hiperparámetros para obtener el mejor rendimiento de la red neuronal. En la metodología de los capítulos correspondientes también se indicarán los valores explorados de estos dos hiperparámetros.

3.1.2. Naïve Bayes

El teorema de Bayes se utiliza para estimar la probabilidad de ocurrencia de un evento, a partir de las probabilidades conocidas de algunas condiciones asociadas a ese evento. El clasificador Naïve Bayes se basa en este teorema y asume la independencia entre los predictores o variables independientes, siendo un modelo fácil de construir [Mitchell 1997]. A pesar de su simplicidad, es utilizado ampliamente, ya que a menudo ofrece un rendimiento sorprendentemente bueno, mejorando en ocasiones los resultados de clasificación obtenidos con métodos mucho más sofisticados.

Dada una muestra X compuesta por un conjunto de características o predictores tal que $X = \{x_1, x_2, \dots, x_n\}$ y un conjunto de clases \mathcal{C} tales que $c \in \mathcal{C}$, el teorema de Bayes proporciona un método para calcular la probabilidad *a posteriori* de la clase c a la que pertenece la muestra X . La ecuación para la descomposición de la probabilidad *a posteriori* de una clase c con la asunción Naïve Bayes ante un conjunto de características X se define según la siguiente expresión.

$$P(c|X) \approx \prod_{i=1}^n P(c|x_i) = \frac{\prod_{i=1}^n P(x_i|c) \cdot P(c)}{P(x_i)} \quad (3.19)$$

El clasificador Naïve Bayes asume que el efecto del valor de un predictor x_i en una clase c es independiente de los valores de otros predictores. Esta asunción se llama independencia condicional de la clase y permite que el clasificador se construya basándose únicamente en el numerador del teorema de Bayes. La regla de decisión básica utilizada para clasificar una nueva muestra como clase c dado un conjunto de predictores X se muestra en la siguiente ecuación.

$$\hat{c} = \arg \max_{c \in \mathcal{C}} \prod_{i=1}^n P(x_i|c) \cdot P(c) \quad (3.20)$$

En conclusión, durante la clasificación de una nueva muestra, se escoge de entre todas las clases aquella que maximiza la probabilidad de pertenecer a ella. De este modo, el resultado depende única y exclusivamente del conjunto de datos de entrenamiento, y de la distribución escogida para modelar la probabilidad.

Durante la experimentación de esta tesis, se utilizaron dos tipos de distribuciones para modelar la probabilidad: 1) distribución Gaussiana, y, 2) *Kernel Density Function* (KDF), detallando en los respectivos capítulos los parámetros que se ajustaron.

3.1.3. Support Vector Machines

En *Support Vector Machines* (SVM) [Cristianini 2000], el proceso de AA estima el hiperplano que maximiza el margen entre dos clases en los datos de entrenamiento. El margen se define como la distancia perpendicular mínima entre dos puntos de cada clase al hiperplano separador. Dicho hiperplano se ajusta durante el proceso de aprendizaje, seleccionando de entre los datos de entrenamiento o predictores (X) aquellos vectores que mejor definen el hiperplano, los cuales son llamados vectores soporte o *support vectors*.

En la Figura 3.5 aparece un esquema básico de la idea que persigue SVM con dos dimensiones o características. Ahí, las muestras aparecen divididas en dos clases: rojo y verde. Aquellas muestras que mejor definen la separación entre clases se utilizan como vectores soporte. El margen muestra la distancia mínima entre dos vectores soporte de distintas clases. En un caso ideal, un plano podría separar dos clases ayudándose de un margen de distancia a dicho plano.

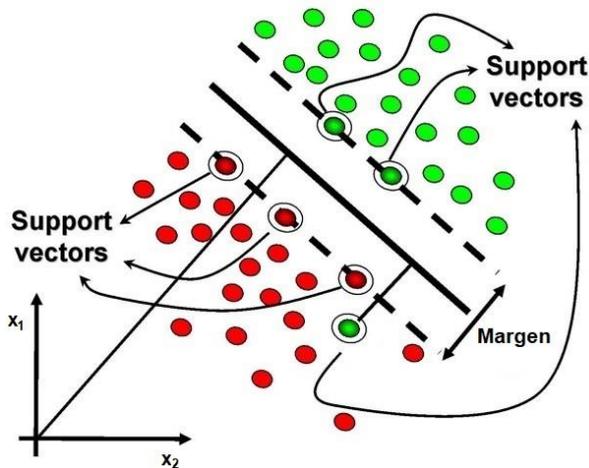


Figura 3.5. Idea básica de hiperplano en SVM en un espacio de dos dimensiones o características (x_1, x_2) . [Winters 2007].

En un problema de clasificación, el número de características o variables independientes utilizadas determina la dimensionalidad D del problema. El hiperplano utilizado en ese espacio para separar dos clases será de dimensión $D - 1$.

Los dos hiperplanos frontera, HP_1 y HP_2 , que definen el margen en SVM se expresan según las ecuaciones 3.21 y 3.22, donde \vec{w} es el vector de pesos que define la dirección del hiperplano, \vec{x} es el vector de características de la muestra de datos que se está clasificando, mientras que b es el término de sesgo o *bias* del hiperplano.

$$HP_1 \rightarrow \vec{w} \cdot \vec{x} + b = -1 \quad (3.21)$$

$$HP_2 \rightarrow \vec{w} \cdot \vec{x} + b = +1 \quad (3.22)$$

Por otra parte, el margen m es la distancia entre los dos hiperplanos frontera anteriores, cuyo valor se determina mediante la siguiente fórmula

$$m = \frac{2}{\|\vec{w}\|} \quad (3.23)$$

Es importante destacar que maximizar el margen de separación entre dos clases, dados unos datos de entrenamiento, tiene como efecto la reducción de la complejidad del clasificador y, por tanto, la optimización de la generalización. Por esta razón, el hiperplano óptimo se determina como aquel que minimiza el error de entrenamiento y, simultáneamente, presenta el margen máximo de separación entre las dos clases. La Figura 3.6 ilustra un ejemplo de hiperplano y margen entre dos clases en SVM, utilizando las dos características x_1 y x_2 .

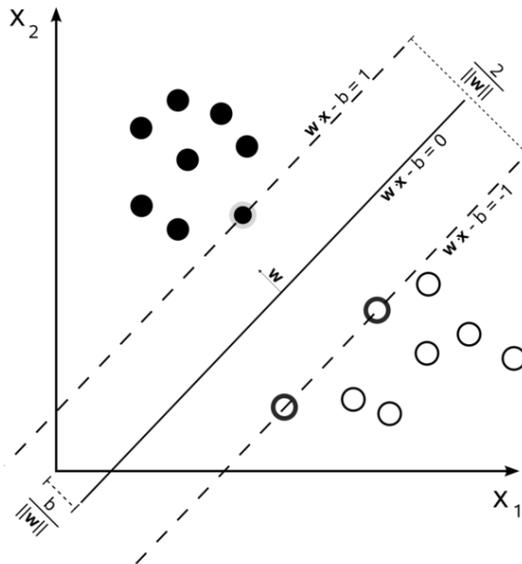


Figura 3.6. Ejemplo de hiperplanos y margen de separación en SVM [Wikimedia Commons, 2023].

Para solucionar el problema de los límites de decisión no linealmente separables, en SVM se emplea una función *Kernel* (K), que transforma implícitamente los datos a un espacio dimensional superior, permitiendo así la separación lineal de las clases, tal como se muestra en la Figura 3.7.

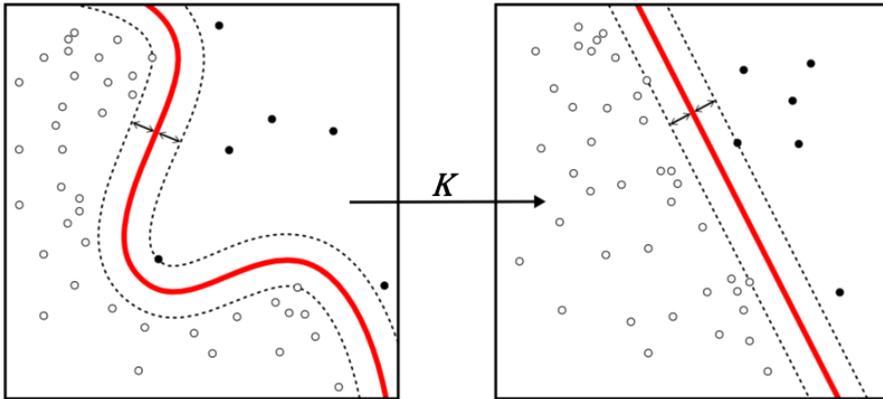


Figura 3.7. Transformación de un espacio no linealmente separable a otro que sí lo es, mediante una función Kernel (K) [Wikimedia Commons, 2023].

La función *Kernel* K puede ser de tipo polinomial, de base radial gaussiana (RBF), o perceptrón sigmoide, entre otras [Gondra 2008]. En la presente tesis se utilizaron únicamente funciones Kernel de tipo RBF, indicando en la metodología de los capítulos correspondientes los valores de los parámetros explorados durante su ajuste.

Después de estimar los parámetros del modelo SVM, incluyendo los vectores soporte y los parámetros de la función *Kernel* seleccionada, se utilizó la ecuación 3.24 para etiquetar nuevas muestras y determinar a qué clase pertenecían. Esta ecuación define la regla de clasificación binaria en SVM, donde se emplea un conjunto de n vectores soporte. Cada elemento s_i representa uno de los vectores soporte utilizados, α_i es el peso asignado a cada vector, b es el *bias* o sesgo, K es la función kernel empleada y \hat{c} es la clase con la que se etiqueta la muestra x . Si el valor del signo de la función es positivo, la muestra se clasifica como miembro del primer grupo; en caso contrario, se clasifica como miembro del segundo grupo.

$$\hat{c} = \text{signo} \left(\sum_{i=1}^n \alpha_i K(s_i, x) + b \right) \quad (3.24)$$

Aunque SVM tiene una gran capacidad de clasificación, una de sus desventajas es que los modelos generados solo pueden utilizarse como clasificadores binarios. Esto se convierte en un inconveniente cuando se abordan problemas de clasificación con más de dos clases o grupos. Para abordar este problema, se han propuesto y evaluado diversas soluciones [Hsu 2002]. Entre ellas, destacan dos enfoques: *One-vs-Rest* y *One-vs-One* [Bishop 2006].

En la estrategia *One-vs-Rest*, se entrena un único clasificador por cada clase, utilizando las muestras de esa clase como positivas (+1) y todas las demás como negativas (-1). Durante la clasificación de nuevas muestras, se utiliza una regla de decisión que tiene en cuenta la salida de todos estos clasificadores para asignar una etiqueta.

En la segunda estrategia, *One-vs-One*, para un problema con C clases, se entrenan un total de $C(C - 1)/2$ clasificadores binarios, cada uno de los cuales recibe las muestras de un par de clases del conjunto de entrenamiento original y debe aprender a distinguir entre ellas. Al clasificar nuevas muestras, se aplica un esquema de votación para asignar la clase que obtuvo más valores positivos (+1) entre todos los $C(C - 1)/2$ clasificadores.

En esta tesis, la estrategia de clasificación adoptada para problemas de más de una clase cuando utilizamos SVM fue *One-vs-One*, mientras que *One-vs-Rest* se empleó con el resto de clasificadores en determinadas circunstancias. En cada capítulo se detalla los parámetros y estrategias seguidos al respecto.

3.1.4. Regresión logística

La regresión logística (RL) es una técnica estadística que se utiliza para modelar la relación entre una variable binaria dependiente y una o más variables independientes continuas o categóricas. Aunque fue descrita hace muchos años, sigue empleándose en diversos campos, como la epidemiología, la biología, la economía, la psicología, la ingeniería y el aprendizaje automático [Cox 1958, Hosmer 2013].

La RL se basa en la función logística comúnmente conocida por *logsig*, que es una función sigmoidea que tiene la apariencia de una curva en forma de S (ver Figura 3.8). Esta función se define matemáticamente mediante la ecuación 3.25, donde z es una variable continua, y e es la base del logaritmo natural.

$$\text{logsig}(z) = \frac{1}{1 + e^{-z}} \quad (3.25)$$

La función *logsig* se utiliza para transformar la salida de una regresión lineal en una probabilidad. En este caso, z es el resultado de la combinación lineal de las variables independientes X_i y sus coeficientes β_i , siguiendo la siguiente expresión:

$$z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n \quad (3.26)$$

De este modo, es posible modelar la relación entre una o más variables independientes y la probabilidad de pertenecer a una categoría o clase específica. Esto es así ya que se transforma el resultado de la combinación lineal de las variables

independientes y sus coeficientes (z) en un valor entre 0 y 1, lo que permite interpretarlo como una probabilidad.

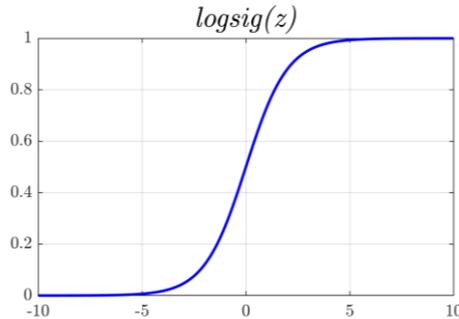


Figura 3.8. Función *logsig* utilizada en las regresiones logísticas

Para ajustar los coeficientes de la RL, se utiliza la función de verosimilitud, que mide la probabilidad de observar los datos de entrenamiento dados los parámetros del modelo. El objetivo del entrenamiento es encontrar los valores óptimos de los coeficientes β_i que maximizan la función de verosimilitud según los datos de entrada.

La selección de un umbral para la salida de la regresión logística es un paso crucial en el proceso de clasificación. El umbral es el valor que se utiliza para separar las clases predichas por el modelo. Por lo general, el umbral se establece en 0,5, lo que significa que, si la probabilidad de pertenecer a la clase positiva es mayor que ese valor, se clasifica como positiva y, en caso contrario, se clasifica como negativa. Sin embargo, en algunos casos, puede ser necesario ajustar el umbral para obtener una clasificación más precisa. En general, la selección del umbral dependerá del equilibrio entre la sensibilidad y la especificidad deseada en el modelo.

Una vez entrenada la RL, y seleccionado el umbral de decisión, se puede usar el modelo para predecir la variable de salida de nuevos datos de entrada que no se han utilizado en el proceso de entrenamiento.

En esta tesis, se detallará la estrategia de selección de umbral empleada en aquellos capítulos donde se emplearon clasificadores basados en RL.

3.2. Sensibilidad y Especificidad

Durante esta tesis se utilizan distintas métricas para analizar el rendimiento de los clasificadores, basadas principalmente en la relación entre sensibilidad y especificidad. Por lo tanto, es necesario definir y establecer previamente estos conceptos utilizados en algunos de los capítulos siguientes.

A partir de la matriz de confusión de la Tabla 3.1, se obtuvieron cuatro valores fundamentales para cada modelo de clasificación binario: Verdaderos Positivos (VP), Verdaderos Negativos (VN), Falsos Negativos (FN), y Falsos Positivos (FP).

Tabla 3.1. Matriz de confusión utilizada para contar aciertos y errores de los clasificadores.

La diagonal de la matriz indica los aciertos, y el resto de elementos los fallos.

		Clase Predicha	
		Arritmia (+)	Normal (-)
Gold Standard	Arritmia (+)	Verdaderos Positivos <i>VP</i>	Falsos Negativos <i>FN</i>
	Normal (-)	Falsos Positivos <i>FP</i>	Verdaderos Negativos <i>VN</i>

La especificidad midió la capacidad de nuestros modelos para identificar muestras negativas, es decir, no pertenecientes a la clase de arritmias. La fórmula que utilizamos para su cálculo aparece en la ecuación 3.27. Valores reducidos indican que nuestro modelo clasifica erróneamente los ritmos normales, mientras que valores elevados muestran una gran eficacia para identificar pacientes sanos [Marinho 2019].

$$\text{Especificidad} = \frac{VN}{VN + FP} \quad (3.27)$$

Por otro lado, la sensibilidad es la proporción de verdaderos positivos y mide la capacidad de nuestros modelos para identificar arritmias. La fórmula que utilizamos para su cálculo aparece en la ecuación 3.28. Una tasa de sensibilidad baja indica que nuestro clasificador está identificando erróneamente la arritmia cardiaca en cuestión, mientras que valores elevados muestran una gran eficacia para identificar patologías.

$$\text{Sensibilidad} = \frac{VP}{VP + FN} \quad (3.28)$$

Es importante destacar que es necesario un equilibrio adecuado entre sensibilidad y especificidad. Por un lado, una falsa alarma podría causar estrés y traumas innecesarios en los pacientes, mientras que una sensibilidad muy baja podría dejar pasar patologías relevantes. En la metodología de los distintos capítulos se indican las estrategias a seguir para balancear estas dos medidas, que se basan principalmente en el uso de la media geométrica (G) entre ambos indicadores, definida como

$$G = \sqrt{\text{Sensibilidad} \times \text{Especificidad}} \quad (3.29)$$

Una de las ventajas de la métrica G es que proporciona una medida de precisión fácil y comprensible que no se ve comprometida por la prevalencia de la enfermedad.

Capítulo 4

DetECCIÓN DE FIBRILACIÓN AURICULAR EN UNA DERIVACIÓN ÚNICA

4.1. Introducción

La fibrilación auricular (FA) destaca entre las enfermedades cardiovasculares al ser el tipo de arritmia sostenida más común en adultos [Hindricks 2021, Chugh 2014], con la tendencia de crecimiento más significativa entre la población, sobre todo en personas de edad avanzada o con trastornos de obesidad [Wang 2004, Wheeldon 1995]. Además, ocasionalmente es difícil de diagnosticar debido a la posible ausencia de síntomas y/o comportamientos paroxísticos [Naccarelli 2009].

Por otro lado, los dispositivos portátiles o *wearables* están ganando interés en la monitorización de pacientes, tanto en ámbitos clínicos como de investigación [Kumari 2017]. Sin embargo, los métodos automáticos para detectar de forma precisa la FA utilizando señales de ECG obtenidas con dispositivos portátiles todavía presenta desafíos, especialmente si se consideran también otros ritmos normales o patológicos [Hermans 2022, Duncker 2021].

En consecuencia, es necesario diseñar e implementar metodologías adecuadas capaces de detectar la FA en registros de ECG de una única derivación que sean computacionalmente apropiadas para dispositivos móviles o portátiles.

4.1.1. Objetivos

El objetivo principal de este estudio es desarrollar y validar una metodología robusta y de bajo coste computacional capaz de identificar la FA en registros ECG de una única derivación. Las señales que utilizamos fueron adquiridas mediante un pequeño dispositivo comercial conectado a un teléfono móvil. Por lo tanto, los

modelos de clasificación resultantes podrían ser candidatos apropiados para implantarse en sistemas de gestión de pacientes que utilicen dispositivos portátiles para la monitorización cardíaca.

Los objetivos detallados de este estudio son los siguientes:

- Definir y extraer automáticamente características derivadas de la actividad ventricular de las señales ECG disponibles.
- Definir, validar y analizar procesos de selección de características.
- Entrenar y evaluar el rendimiento de distintos modelos de clasificación al etiquetar registros cortos de ECG de una única derivación entre cuatro grupos diferentes: ritmo sinusal normal, fibrilación auricular, otros ritmos y registro ruidoso.

4.2. Materiales

Para validar el rendimiento del enfoque propuesto, utilizamos la base de datos donada por la empresa AliveCor en el contexto del *PhysioNet/Computing in Cardiology Challenge 2017* [Clifford 2017], que contenía muestras de ECG de una única derivación.

En este estudio utilizamos únicamente los 8.528 registros de ECG proporcionados como conjunto de entrenamiento público, los cuales tenían duraciones de entre 9 y 61 segundos. La Tabla 4.1 muestra las cuatro categorías o clases de ritmos cardíacos a detectar, el número y la duración de las muestras del conjunto de datos de entrenamiento. Cada etiqueta fue revisada por cardiólogos expertos.

Tabla 4.1. Nombre, abreviatura, número y duración de los registros ECG utilizados.

Nombre (Abreviatura)	Nº de Muestras	Duración (segundos)
		Media±Desv. [Mín. - Máx.]
Ritmo Sinusal Normal (RSN)	5076 (59,5%)	32,1±10,0 [9,0 - 61,0]
Fibrilación Auricular (FA)	758 (8,9%)	32,3±12,3 [10,0 - 60,2]
Otros Ritmos (OR)	2415 (28,3%)	34,3±11,8 [9,1 - 60,9]
Registro Ruidoso (RD)	279 (3,3%)	24,4±10,4 [9,4 - 60,0]
Total	8528 (100%)	32,5±10,9 [9,0 - 61,0]

Cada registro ECG se adquirió por una persona en una de las tres generaciones del dispositivo Kardia Mobile™, de la empresa AliveCor (www.alivecor.com). Cada usuario presionó o sostuvo en sus manos un electrodo con el fin de obtener un ECG equivalente a la derivación I. La Figura 4.1 ilustra el modo de uso del dispositivo.

Cabe destacar que muchos de los ECG de la base de datos aparecen invertidos ya que no se requiere que se presionen o sostengan los electrodos en una orientación particular concreta.



Figura 4.1. Dispositivo Kardia Mobile™ (AliveCor) conectado a la aplicación móvil Kardia App, adquiriendo un ECG equivalente a la derivación I.

El dispositivo registra el ECG durante un periodo de tiempo medio de 30 segundos después de algunas comprobaciones previas de la calidad de la señal. A continuación, transmite acústicamente los datos a un *smartphone* o tableta, donde son digitalizados en tiempo real a 44.1 kHz y una resolución de 24 bits, utilizando para ello demodulación por software. Finalmente, la señal ECG se almacena en ficheros muestreados a una frecuencia de 300 Hz, con una resolución de 16 bits y un rango dinámico de ± 5 mV, aplicando previamente un filtro pasa-banda de 0,5-40 Hz.

La Figura 4.2 muestra un ejemplo de registros ECG de cada una de las clases a detectar adquiridos por el dispositivo.

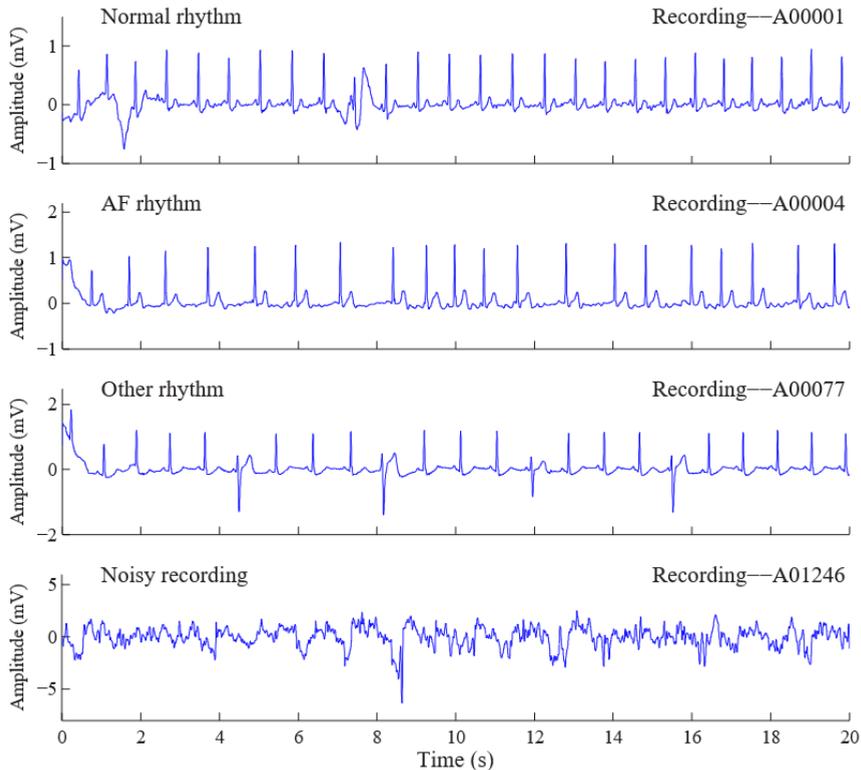


Figura 4.2. Ejemplos de ECG de la base de datos utilizada: ritmo normal, fibrilación auricular, otros ritmos anómalos y registro ruidoso [Clifford 2017].

4.3. Métodos

En esta sección describimos en primer lugar los criterios de evaluación para los modelos de clasificación utilizados. Seguidamente se detalla el preprocesado de señal aplicado a los registros ECG, y se detalla el proceso de extracción de características empleado. A continuación, se describe el preprocesado de datos, la selección de características utilizada y la metodología de validación de los modelos de clasificación.

El entorno de trabajo donde realizamos todos estos pasos fue MATLAB (R2017a, The MathWorks); el código desarrollado está disponible en <https://physionet.org/content/challenge-2017>. Por otra parte, el hardware utilizado en este estudio fue un ordenador con procesador Intel® Core™ i7-4790 @ 3.6 GHz, 16 GB de memoria RAM, bajo el sistema operativo Windows 10 Enterprise (64 bits). Además, se utilizó la *toolbox* externa *DeepLearnToolbox* [Palm 2012], donde se encuentran funcionalidades para el entrenamiento de redes neuronales artificiales.

Finalmente, presentaremos los costes computacionales de los procesos de extracción de características y clasificación de los modelos desarrollados en términos de número de muestras procesadas por segundo, así como de milisegundos necesarios para procesar una única muestra.

4.3.1. Criterios de evaluación

Los porcentajes de aparición de cada una de las clases varían significativamente (ver Tabla 4.1), lo que indica que la base de datos empleada está desequilibrada [Japkowicz 2002]. Estos porcentajes permitirían que un clasificador que siempre asignara la clase más frecuente (ritmo normal) a nuevas muestras tuviera una precisión cercana al 60% con las muestras que disponíamos. Sin embargo, la sensibilidad real ante ritmos patológicos sería nula.

Para evitar dicho sesgo, el criterio de evaluación escogido por los organizadores del *PhysioNet/Computing in Cardiology Challenge 2017* [Clifford 2017] fue la medida *F-Score* (F_1), ponderando así la precisión y la exhaustividad (*precision and recall* en inglés) [Sasaki 2007]. F_1 es una función definida en el rango (0, 1), donde valores elevados también indican una relación equilibrada entre precisión y exhaustividad. El criterio de evaluación de este estudio es el mismo que el utilizado en la competición de 2017: un valor F_1 promedio de entre las distintas clases, descrito a continuación.

En primer lugar, la Tabla 4.2 muestra las reglas de conteo para cada uno de los aciertos y errores de clasificación entre clases.

Tabla 4.2. Reglas de conteo de aciertos y errores entre las clases a etiquetar.

La diagonal de la matriz indica los aciertos, y el resto de elementos los fallos.

		Clase Predicha				Total
		RSN	FA	OR	RD	
Gold Standard	RSN	Nn	Na	No	Nd	$\sum N$
	FA	An	Aa	Ao	Ad	$\sum A$
	OR	On	Oa	Oo	Od	$\sum O$
	RD	Dn	Da	Do	Dd	$\sum D$
	Total	$\sum n$	$\sum a$	$\sum o$	$\sum d$	

A continuación, a partir de las reglas de conteo anteriores, se definen las puntuaciones individuales de *F-Score* para cada una de las cuatro clases a identificar mediante las siguientes expresiones:

Puntuación individual para Ritmo Sinusal Normal:

$$F_{rsn} = \frac{2 \times Nn}{\sum N + \sum n} \quad (4.1)$$

Puntuación individual para Fibrilación Auricular:

$$F_{fa} = \frac{2 \times Aa}{\sum A + \sum a} \quad (4.2)$$

Puntuación individual para Otros Ritmos:

$$F_{or} = \frac{2 \times Oo}{\sum O + \sum o} \quad (4.3)$$

Puntuación individual para registros ruidosos:

$$F_{rd} = \frac{2 \times Dd}{\sum D + \sum d} \quad (4.4)$$

El rendimiento de clasificación medio viene definido por el valor de la métrica \bar{F}_1 , la cual representa el promedio entre los valores F_{rsn} , F_{fa} y F_{or} , definido como:

$$\bar{F}_1 = \frac{F_{rsn} + F_{fa} + F_{or}}{3} \quad (4.5)$$

Aunque F_{rd} no se incluye en este promedio, sí que afecta en los valores individuales de *F-Score* para cada clase.

Finalmente, se realizó una validación cruzada de diez conjuntos o pliegues, utilizando las 8.528 muestras disponibles, para evaluar el rendimiento final de los modelos. Esto es, para cada uno de los diez pliegues ($K=10$), se entrenaron los modelos con todas las combinaciones posibles de $K - 1$ particiones, donde el rendimiento en cada ronda con la partición restante se identifica mediante el término $\bar{F}_1(k)$. El promedio de los K valores se define como:

$$F_1 = \frac{1}{K} \sum_{k=1}^K \bar{F}_1(k) \mid K = 10 \quad (4.6)$$

Esta última medida F_1 se utilizó para evaluar los modelos de clasificación del presente estudio. En la sección de resultados mostramos los mejores valores de F_1 obtenidos durante la validación cruzada, además de los valores medios (media±desviación estándar) del F -Score para cada una de las clases (F_{rsn} , F_{fa} , F_{or} , F_{rd}) en los diferentes modelos de clasificación empleados en la misma validación cruzada.

4.3.2. Preprocesado de señal

Los registros ECG de la base de datos estaban previamente filtrados en la banda entre 0,5 Hz y 40 Hz. Por lo tanto, la única etapa de preprocesado de señal que realizamos fue un filtrado de artefactos. Para ello, utilizamos una ventana deslizante de 0,5 segundos, calculando en cada segmento los valores aberrantes máximos y mínimos. Finalmente, anulamos aquellos tramos cuya ventana contenía valores atípicos, tal como se ilustra en la Figura 4.3.

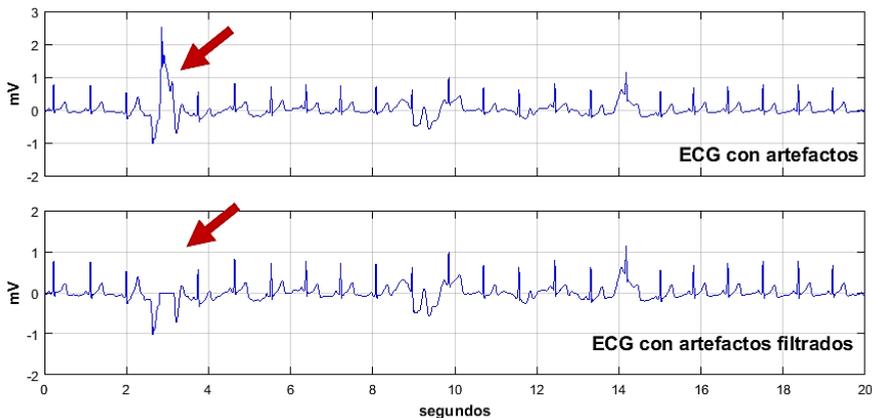


Figura 4.3. Proceso automático de filtrado de valores aberrantes realizado en las señales ECG.

4.3.3. Extracción de características

Extrajimos automáticamente 81 características de la señal ECG derivadas de la actividad ventricular. Se evitaron las características relacionadas con la actividad auricular ya que su señal puede verse enmascarada por ruido y/o artefactos en registros de una derivación. Estos ruidos y artefactos son especialmente relevantes en dispositivos portables de uso personal por varios motivos: 1) por ser diseños de bajo coste, y por tanto con peor relación señal-ruido; 2) por la alta frecuencia de artefactos y fluctuaciones de la línea base a causa de movimientos y 3) por posibles usos inadecuados, debido a que los usuarios de los dispositivos son, en general,

inexpertos y sin la formación adecuada que sí posee el personal clínico responsable del manejo de un equipo electrocardiográfico de uso hospitalario.

En primer lugar, aplicamos un detector de ondas QRS basado en el algoritmo de Pan y Tompkins y la primera derivada del ECG [Pan 1985]. Filtramos las detecciones de aquellas ondas QRS que no superasen un valor medio de 0,75 en el coeficiente de correlación con todas las demás detecciones, utilizando una ventana de 200 ms. De este modo evitamos falsos positivos, eliminando la mayoría de detecciones en latidos ectópicos y artefactos. La Figura 4.4 ilustra el funcionamiento de nuestro detector de ondas QRS con registros ECG etiquetados con los ritmos cardiacos a identificar de la base de datos utilizada, donde los puntos rojos indican detecciones de los picos R.

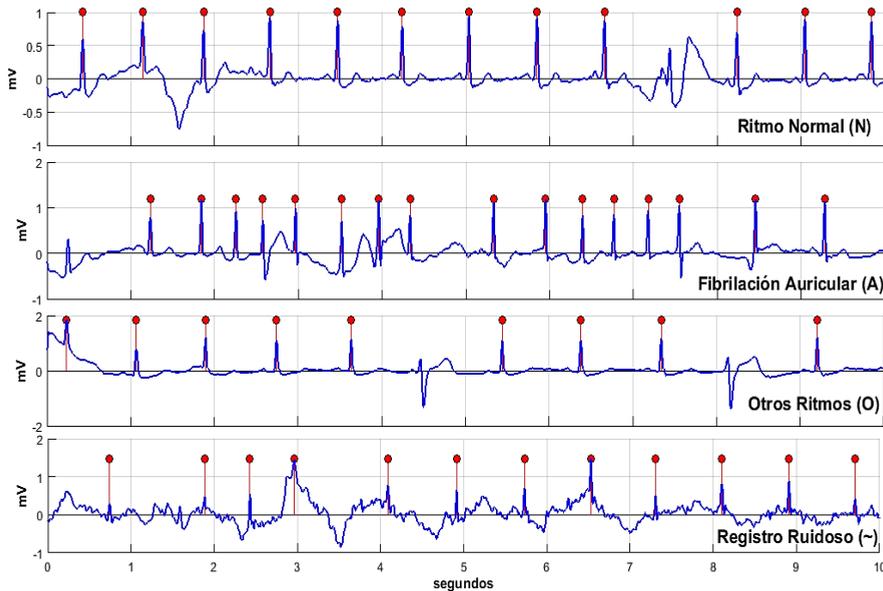


Figura 4.4. Ejemplos de detección de complejos QRS en registros ECG de la base de datos utilizada.

A continuación, extrajimos la secuencia RR a partir de los picos R detectados, donde cada elemento i -ésimo de la secuencia viene definido por la siguiente expresión:

$$RR_i = R_{i+1} - R_i \quad (4.7)$$

La secuencia RR representa la derivada de la secuencia temporal de los picos R detectados, donde cada uno de sus elementos RR_i representan la diferencia temporal (en milisegundos) entre dos latidos consecutivos. A continuación, filtramos

los valores atípicos en RR , y obtuvimos su primera ($RRd1$) y segunda ($RRd2$) derivada, definidas como

$$RRd1_i = RR_{i+1} - RR_i \quad (4.8)$$

$$RRd2_i = RRd1_{i+1} - RRd1_i \quad (4.9)$$

Nótese que estas secuencias (RR , $RRd1$, $RRd2$) son utilizadas frecuentemente para el análisis de la variabilidad del ritmo cardiaco en registros de larga duración que implican horas o incluso días [Tsuji 1996]. No obstante, la duración de nuestros registros en comparación es corta, con duraciones que oscilan entre los 10 y 60 segundos.

Utilizando las marcas de ondas QRS y las secuencias RR , agrupamos en siete categorías diferentes las 81 características de señal extraídas de la única derivación ECG disponible, las cuales se detallan en la Tabla 4.3. El buen funcionamiento del detector de ondas QRS está estrechamente relacionado al buen rendimiento de la clasificación final ya que todas las características extraídas dependen de tales marcas fiduciales. Puede encontrarse más información sobre la relevancia clínica de estas características en [Shaffer 2017].

Tabla 4.3. Detalle de las siete categorías de características extraídas del ECG. Número y descripción de características en cada categoría.

Categoría	Nº	Descripción
1. Estadísticas RR	9	Estadísticas básicas sobre las secuencias RR , $RRd1$ y $RRd2$ seleccionando sólo aquellas que presentaron diferencias significativas entre al menos dos clases: media (RR , $RRd1$), desviación estándar (RR , $RRd1$, $RRd2$), curtosis (RR , $RRd1$, $RRd2$) y asimetría (RR).
2. Basadas en pNNx	4	Características complejas basadas en la secuencia $RRd1$: Media cuadrática de diferencias sucesivas (RMSSD). pNN25, pNN50, pNN75, donde pNNx indica el porcentaje de intervalos entre latidos normales que exceden los 25, 50 y 75 ms [Mietus 2002].
3. Estadísticas de Ventana Deslizante	44	Usando una ventana deslizante en la señal de 8 segundos en intervalos consecutivos de 2 segundos extrajimos para cada tramo de señal las características de las categorías 1 y 2, excepto la asimetría del RR , y la media de

Categoría	Nº	Descripción
		RRd1 (11 elementos). Una vez obtenida la matriz de valores para cada intervalo de señal, extrajimos las siguientes cuatro estadísticas de cada columna: mínimo, máximo, media y desviación estándar (44 elementos).
4. Gráfica de Poincaré	5	Características basadas en la gráfica de Poincaré utilizando la secuencia <i>RRd1</i> : Valores máximos, mínimos, media, desviación estándar de las distancias entre todos los puntos de la gráfica, así como la diferencia absoluta entre los valores de distancia máxima y mínima.
5. Gráfica de Lorenz	8	Características basadas en la gráfica de Lorenz utilizando la secuencia <i>RRd2</i> : Variabilidad angular: <i>Vector Angular Index (VAI)</i> y <i>Vector Length Index (VLI)</i> [Ruan 2011]. Media y desviación de las distancias de los puntos al origen, de las distancias entre dos puntos consecutivos, y de las distancias entre tres puntos consecutivos en <i>RRd2</i> .
6. Características complejas	3	Complejidad de Lempel-Ziv de la serie temporal <i>RR</i> después de la binarización utilizando la mediana como umbral. Entropía de Shannon de la secuencia <i>RR</i> . Ratio entre el número de distintos patrones QRS encontrados y el número total de ondas R detectadas (PPC).
7. Distancias a los vectores $pNNx$ medios	8	Dos medidas de distancia entre el vector $pNNx$ medio [Mietus 2002] de cada clase ($pNNx(\text{RSN})$, $pNNx(\text{FA})$, $pNNx(\text{OR})$, $pNNx(\text{RD})$) y el vector $pNNx(\hat{i})$ correspondiente a cada registro ECG: <i>Dynamic Time Warping</i> a cada clase ($dtw\text{RSN}$, $dtw\text{FA}$, $dtw\text{OR}$, $dtw\text{RD}$) [Jiménez 2017a]. Error cuadrático a cada clase ($ec\text{RSN}$, $ec\text{FA}$, $ec\text{OR}$, $ec\text{RD}$).
Número de características totales: 81		

A continuación, se describe detalladamente la forma de obtener los valores de la última categoría de características basadas en distancias a los vectores $pNNx$ medios.

En primer lugar, es necesario definir el concepto de vector $pNNx$, el cual no debe confundirse con el de los valores de los indicadores $pNN25$, $pNN50$, $pNN75$, aunque esté relacionado con ellos. Denotamos con x al umbral (en ms) utilizado para indicar el porcentaje de latidos normales consecutivos cuya diferencia temporal excede dicho umbral. Por ejemplo, en $pNN25$, $pNN50$, $pNN75$, el umbral x entre latidos normales es de 25, 50 y 75 ms respectivamente. Es posible obtener un vector $pNNx$ de un registro ECG llamado i si variamos x en un rango determinado. De este modo, para cada valor del umbral obtendremos el porcentaje de latidos consecutivos que difieren en más de x milisegundos. En nuestro caso utilizamos un rango de umbrales entre 10 y 150 ms, definiendo dicho vector como

$$pNNx(i) = \frac{\text{tamaño}(RRd1 > x)}{\text{tamaño}(RRd1)} \quad \forall x \in [10, 150] \quad 4.10$$

Seguidamente, para cada una de las clases a detectar, se definió un vector $pNNx$ medio mediante las siguientes expresiones, donde n_{RSN} , n_{FA} , n_{OR} y n_{RD} indican el número de muestras de cada clase respectivamente.

$$pNNx(RSN) = \frac{\sum pNNx(i)}{n_{RSN}} \quad \forall i \in RSN \quad 4.11$$

$$pNNx(FA) = \frac{\sum pNNx(i)}{n_{FA}} \quad \forall i \in FA \quad 4.12$$

$$pNNx(OR) = \frac{\sum pNNx(i)}{n_{OR}} \quad \forall i \in OR \quad 4.13$$

$$pNNx(RD) = \frac{\sum pNNx(i)}{n_{RD}} \quad \forall i \in RD \quad 4.14$$

Estos vectores necesitaron calcularse mediante un procesamiento previo de la base de datos, guardando a continuación sus valores para poder reutilizarlos en cada nueva extracción de características.

Usando los vectores $pNNx(RSN)$, $pNNx(FA)$, $pNNx(OR)$ y $pNNx(RD)$, se propusieron las siguientes ocho características basadas en la idea de distancia y similitud entre dos secuencias o vectores. En primer lugar, se utilizó la métrica *Dynamic Time Warping (DTW)* [Sakoe 1978] para calcular una distancia entre el vector $pNNx$ generado con cada muestra i , y el vector $pNNx$ medio de cada clase. Estas nuevas características de distancia dtw se definen como:

$$dtwRSN(i) = DTW(pNNx(RSN), pNNx(i)) \quad (4.15)$$

$$dtwFA(i) = DTW(pNNx(FA), pNNx(i)) \quad (4.16)$$

$$dtwOR(i) = DTW(pNNx(OR), pNNx(i)) \quad (4.17)$$

$$dtwRD(i) = DTW(pNNx(RD), pNNx(i)) \quad (4.18)$$

Del mismo modo, también se utilizó como medida de distancia el error cuadrático (EC) entre los vectores $pNNx(i)$ y los cuatro vectores medios calculados previamente, extrayendo las siguientes cuatro características definidas como:

$$ecRSN(i) = \sum (pNNx(RSN) - pNNx(i))^2 \quad (4.19)$$

$$ecFA(i) = \sum (pNNx(FA) - pNNx(i))^2 \quad (4.20)$$

$$ecOR(i) = \sum (pNNx(OR) - pNNx(i))^2 \quad (4.21)$$

$$ecRD(i) = \sum (pNNx(RD) - pNNx(i))^2 \quad (4.22)$$

Puesto que esta categoría de características basadas en distancias (DTW y EC) a los vectores $pNNx$ medios es de especial interés en el presente estudio al haber sido definida aquí, realizamos un análisis estadístico de los valores obtenidos que se presenta en los resultados.

En último lugar, extrayendo las 81 características descritas de cada uno de los 8.528 registros ECG disponibles, creamos el conjunto de datos para el entrenamiento de los modelos de clasificación.

4.3.4. Preprocesado del conjunto de datos

Preprocesamos el conjunto de datos formado por las características extraídas utilizando tres etapas consecutivas: corrección de valores atípicos, eliminación de muestras con errores y normalización de los datos.

En primer lugar, para cada característica, los valores atípicos cuyo valor absoluto superaba 3 veces la desviación estándar más la mediana, fueron reemplazados por este mismo límite en su correspondiente valor inferior o superior.

A continuación, eliminamos las muestras si alguna de sus características presentaba un error de extracción identificado por un valor NaN (*Not a Number*).

Siguiendo esta regla, eliminamos un total de 32 muestras (5 de ritmo normal, 7 de FA, 2 de otros ritmos y 18 de registros ruidosos) obteniendo un conjunto de datos final compuesto por 8496 muestras. En este sentido, puesto que más de la mitad de las muestras eliminadas correspondían a la clase de registros ruidosos, establecimos como regla de clasificación asignar la etiqueta de registro ruidoso a aquellas muestras donde existiera algún valor *NaN* entre sus características extraídas.

Posteriormente, para cada pliegue de la validación cruzada se obtuvo la media (μ) y la desviación estándar (σ) de cada característica utilizando las muestras de entrenamiento. Con estos valores, aplicamos una normalización *z-score* a todas las muestras, tanto a las utilizadas para entrenamiento como a las de validación, donde x representa el valor de entrada y z el valor normalizado obtenido, determinado por la siguiente expresión:

$$z = \frac{x - \mu}{\sigma} \quad (4.23)$$

4.3.5. Selección de características

Para la selección de características utilizamos dos procesos de tipo envoltorio o *wrapper* diferentes: *Forward-Selection (FS)* y *Backward-Elimination (BE)* definidos inicialmente por [Kohavi 1997].

En ambas aproximaciones, la idea básica es añadir (*FS*) o eliminar (*BE*) características dependiendo de la dirección de la búsqueda, y evaluar una función de coste que indica si dicha operación mejora o empeora el rendimiento de clasificación. En *FS* se comienza con un conjunto de características vacío, mientras que en *BE* se empieza con todo el conjunto de características. Este proceso se repite seleccionando o eliminando aquella característica que mayor beneficio proporciona en la función de coste, hasta que no se observan mejoras tras añadir o eliminar alguna característica más.

En este estudio, la función de coste evaluada en ambos *wrappers* fue la métrica F_1 aplicada a los resultados de clasificación de un modelo de tipo *Feed Forward Neural Network (FFNN)*. El objetivo aquí no es encontrar el mejor rendimiento de clasificación, sino conjuntos de características reducidos que, combinados, sean viables para proporcionar buenos resultados de clasificación durante entrenamientos más exhaustivos. Por lo tanto, es imprescindible evitar el sobreajuste en esta etapa. Para ello, utilizamos una arquitectura de red de 1 capa oculta, un número reducido de unidades (128) y 75 *epochs* de entrenamiento. Se utilizó una *hold-out validation*, donde el 50% de las muestras se destinaron al entrenamiento del modelo, mientras que con el 50% restante se obtuvo el rendimiento de clasificación mediante la medida F_1 . En un paso inicial, ajustamos los valores de los siguientes parámetros de la FFNN: velocidad de aprendizaje o *learning rate* (α) e impulso o *momentum* (β).

Con los parámetros de la red establecidos (128 unidades, y los valores α y β ajustados), y usando la misma validación *hold-out*, se utilizó el valor F_1 en la función de coste en ambas estrategias de selección de características (FS y BE), para ser optimizado. A los conjuntos de características seleccionadas en estos procesos, los identificaremos por λ_{FS} o λ_{BE} respectivamente según el método de selección empleado.

En la aproximación BE relajamos el criterio de parada del siguiente modo: si durante la iteración i el valor de F_1 no aumenta respecto a la anterior ($i - 1$), seleccionamos como característica a eliminar aquella que maximizaba F_1 durante la iteración i si y solo si $F_{1i} > (F_{1i-1} - 0,05)$. En caso contrario, el proceso terminaba. De este modo, conseguimos conjuntos de características más reducidos con rendimientos similares e incluso mejores que en iteraciones donde el conjunto era mayor.

Las dos aproximaciones de selección de características utilizadas son subóptimas, pudiendo existir otro conjunto que proporcione mejor rendimiento. No obstante, en la práctica suelen proporcionar buenas soluciones en espacios de búsqueda que sufren explosión combinatoria, como es el nuestro, donde éste es del orden $81!$, siendo 81 el número de características con el que trabajamos.

4.3.6. Entrenamiento y validación de Redes Neuronales

Utilizamos redes neuronales FFNN con arquitectura de una capa oculta. Para cada conjunto de características seleccionadas (λ_{FS} y λ_{BE}), realizamos una búsqueda de la combinación de parámetros que mejores resultados de clasificación ofrecía. El número de unidades ($\#u$) explorado de la capa oculta fue de 64, 128, 256, 512 y 1024 respectivamente. Respecto al *learning rate* (α), se probaron los valores de 0,5, 0,6, 0,7, 0,8, 0,9 y 1,0 respectivamente. Finalmente, en cuanto al *momentum* (β), se utilizaron valores de 0,0, 0,1, 0,2, 0,3, 0,4 y 0,5 respectivamente.

Para actualizar los pesos de las redes neuronales se empleó la optimización *Levenberg-Marquardt backpropagation* por descenso de gradiente. Además, se estableció un máximo de 100 *epochs* y un incremento mínimo del gradiente de 0,001 entre dos *epochs* consecutivas como condiciones de parada del entrenamiento.

Finalmente, la función ReLU se utilizó como función de activación, excepto para las unidades de la última capa de la red donde usamos la función *softmax* [Bishop 2006]. Asimismo, en cada *epoch* se actualizó únicamente la mitad de los pesos de las unidades al usar un valor de *dropout* de 0,5.

Además de un análisis del rendimiento de clasificación en el espacio de búsqueda, presentamos los resultados de los modelos que obtuvieron la puntuación más alta en cada conjunto de características seleccionadas, identificados por $FFNN\lambda_{FS}$ y $FFNN\lambda_{BE}$ respectivamente.

4.3.7. Entrenamiento y validación de Naïve Bayes

Para los modelos de tipo Naïve Bayes, realizamos una búsqueda en los siguientes parámetros, escogiendo aquel con mejor rendimiento de clasificación en la validación cruzada: probabilidad *a priori* de las clases y distribución de datos.

Respecto las probabilidades *a priori* de las clases, probamos dos configuraciones: (1) empíricas, y, (2) uniformes. En el primer caso, éstas son determinadas empíricamente a partir del conjunto de datos de entrenamiento; es decir, ante un conjunto de datos de dos clases, donde una de ellas corresponde al 75% de las muestras, los valores de probabilidades *a priori* corresponden a 0,75 y 0,25 respectivamente para cada clase. En el segundo caso, para el mismo problema de clasificación binaria, utilizar probabilidades uniformes daría valores de probabilidad *a priori* de 0,5 a cada clase. En la práctica, utilizar probabilidades uniformes significa eliminar este término de la ecuación del modelo de Naïve Bayes al no alterar en ningún caso la clasificación resultante.

Respecto la distribución de los datos, utilizada para modelar la probabilidad de cada una de las características, probamos también dos configuraciones: 1) distribución Gaussiana, y, 2) *Kernel Density Function (KDF)*.

La distribución Gaussiana está definida por una media (μ) y una desviación típica (σ), donde la probabilidad de aparición del valor x se modela mediante la siguiente expresión:

$$G_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (4.24)$$

Por otro lado, la distribución de datos de tipo *KDF* se define como:

$$kdf(x) = \frac{1}{nW} \sum_{i=1}^n K\left(\frac{x-x_i}{W}\right) \quad (4.25)$$

donde K es la función *Kernel* a utilizar, n el número de muestras de entrenamiento y W el parámetro libre de suavizado conocido por *width*, *bandwidth* o *kernel smoother window width*. Es posible indicar un valor W para cada una de las clases, intentando así mejorar el rendimiento del clasificador Naïve Bayes.

En este estudio, para las distribuciones de probabilidad *KDF*, se utilizaron todas las posibles combinaciones entre los valores 0,005, 0,020, 0,035 y 0,050 para el *bandwidth* de cada clase (W_{rsn} , W_{fa} , W_{or} , W_{rd}). Este rango de valores W fue seleccionado empíricamente. Por otra parte, utilizamos una distribución normal con media 0 y desviación estándar 1 como función *Kernel*, definida como:

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-0.5x^2} \quad (4.26)$$

Esta búsqueda se realizó en los dos conjuntos de características seleccionadas, identificando por $NB\lambda_{FS}$ y $NB\lambda_{BE}$ respectivamente los modelos que obtuvieron la puntuación más alta en cada conjunto.

4.3.8. Entrenamiento y validación de *Support Vector Machines*

Los modelos SVM son clasificadores binarios que no necesitan de la selección de umbral de decisión; esto es así ya que clasifican cada muestra como positiva o negativa dependiendo del lado del hiperplano que le corresponde. Para crear un clasificador de más de dos clases aplicamos la estrategia *One-vs-One*, entrenando seis clasificadores binarios al existir cuatro clases en nuestro conjunto de datos, e implementando posteriormente el correspondiente sistema de votación con los seis modelos SVM binarios.

Por otro lado, en este estudio se utilizaron únicamente *kernels* de base radial gaussiana (RBF), definidos mediante la siguiente expresión:

$$K_{rbf}(s_i, x_j) = \exp\left(\gamma \|s_i - x_j\|^2\right) \quad (4.27)$$

Aquí, la tupla s_i representa un *vector soporte*, x_j la muestra a clasificar y el parámetro libre γ provee un factor de escala para la función.

Inicialmente, ajustamos el valor de γ realizando una búsqueda en el rango [1.0, 3.0] con incrementos de 0,1 respectivamente.

Por otro lado, el parámetro C de los modelos SVM controla la compensación entre errores de entrenamiento y los márgenes rígidos, creando así lo que se denomina un margen blando (*soft margin*), que al mismo tiempo permite y penaliza algunos errores de clasificación [Gondra 2008]. Este parámetro provee de cierta flexibilidad al modelo, representando a la vez el compromiso entre el tamaño del margen y el número de errores de clasificación.

Inicialmente, ajustamos el valor de C probando los valores comprendidos en el rango [0,2, 1,5] con incrementos de 0.1 respectivamente. En un segundo ajuste más fino de los parámetros del modelo, se exploraron valores de C de 1,0, 1,1 y 1,2 respectivamente, junto con valores de γ en el rango [1,3, 1,5] en incrementos sucesivos de 0,02.

Al igual que en los modelos de clasificación anteriores, la búsqueda de parámetros se realizó utilizando los dos conjuntos de características λ_{FS} y λ_{BE} , identificando por $SVM\lambda_{FS}$ y $SVM\lambda_{BE}$ los modelos que obtuvieron la puntuación más alta en cada conjunto.

4.4. Resultados

4.4.1. Análisis de las características basadas en los vectores $pNNx$

En la Tabla 4.4 se muestran los valores obtenidos para cada clase en las características basadas en distancias a los vectores $pNNx$ medios (porcentaje de latidos consecutivos que difieren más de x ms), destacando en negrita aquellos valores donde la distancia menor coincide con la clase donde procede el vector $pNNx$ medio utilizado. En dicho análisis se observa que se cumple esta condición tanto para RSN como para FA en las distancias $dtwRSN$, $dtwFA$, $ecRSN$ y $ecFA$.

La Figura 4.5 muestra los valores de los vectores $pNNx$ medios de cada clase donde las curvas formadas al variar el umbral x prácticamente se estabilizan en todas las clases entre 75 y 100 ms excepto en la FA, donde siempre aparece una tendencia descendente mayor. Por otro lado, la Figura 4.6, muestra el valor medio de cada punto junto con la desviación estándar de los vectores $pNNx$ para las clases RSN y FA. Ahí puede apreciarse el alto grado de separación que se consigue entre estos dos grupos. Además, cabe destacar que utilizar un único valor del vector, como por ejemplo el $pNN50$ o el $pNN75$ medio también permite obtener un buen nivel de división entre RSN y FA.

No obstante, existe un solapamiento de los valores de RSN y FA con los valores de las clases OR y RD. Para ilustrar este comportamiento, las Figura 4.7 y 4.8 presentan los diagramas de caja de las variables $dtwRSN$ y $dtwRD$ respectivamente. En ellas puede observarse como los valores de distancia son menores en las muestras que pertenecen a la clase RSN, tal como cabe esperar. Sin embargo, en ambas características, los valores de la clase FA se solapan principalmente con los valores de registros ruidosos, mientras que RSN se solapa mayoritariamente con la clase OR.

Tabla 4.4. Valores de las características basadas en distancias a los vectores $pNNx$ medios.

	i=RSN	i=FA	i=OR	i=RD
$dtwRSN$	0,97±0,67	2,83±1,04	1,42±1,14	3,19±1,83
$dtwFA$	2,88±0,78	1,06±0,66	2,39±1,01	1,80±1,15
$dtwOR$	1,33±0,48	2,18±0,91	1,41±0,83	2,58±1,67
$dtwRD$	3,17±0,80	1,20±0,69	2,58±1,12	1,68±1,18
$ecRSN$	1,39±3,00	9,08±6,07	3,30±5,58	13,52±16,07
$ecFA$	8,91±3,81	1,55±2,08	6,72±4,52	4,56±8,05
$ecOR$	2,00±1,98	5,59±4,43	2,69±3,66	9,41±13,41
$ecRD$	10,70±4,13	1,91±2,34	7,89±5,23	4,21±7,55

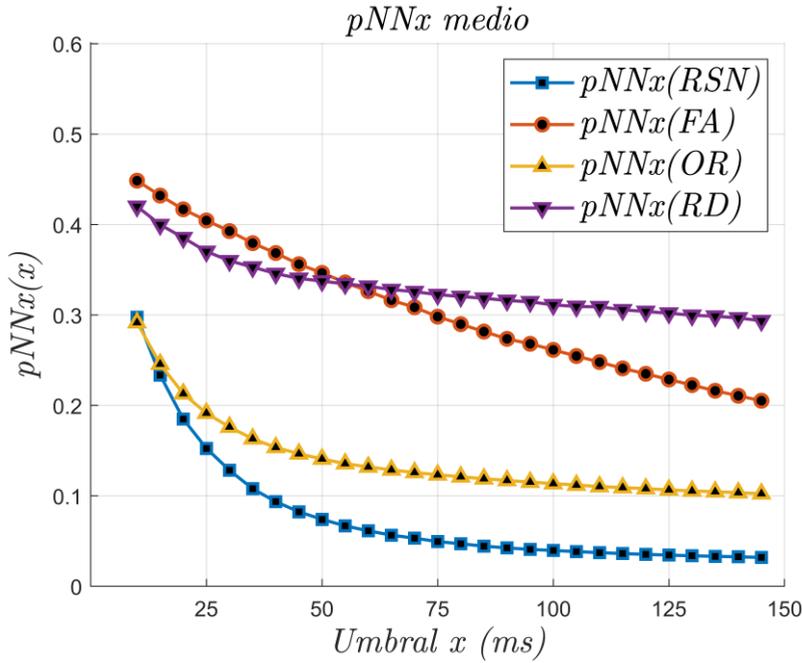


Figura 4.5. Vectores $pNNx$ medios de cada una de las clases (RSN, FA, OR, RD).

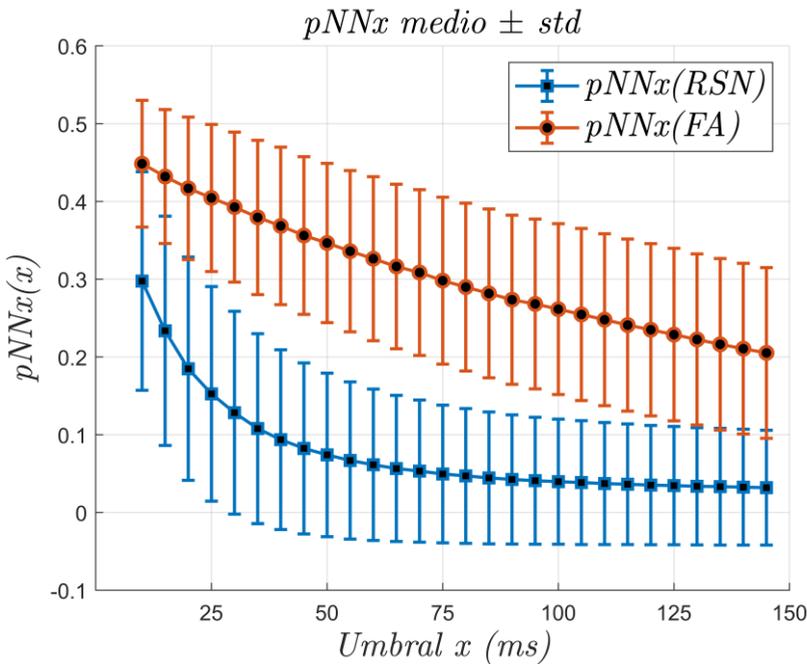


Figura 4.6. Media y desviación de los vectores $pNNx$ para las clases RSN y FA.

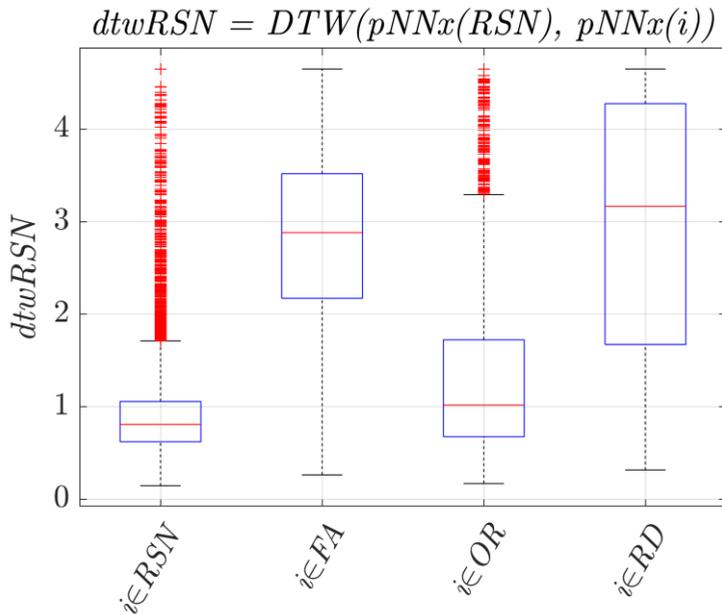


Figura 4.7. Diagramas de caja de los valores de $dtwRSN$ para cada una de las clases.

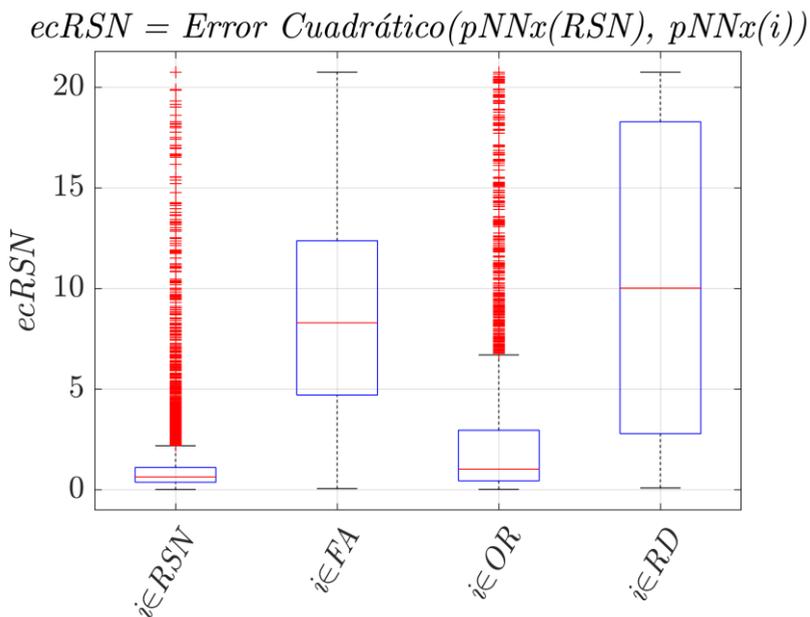


Figura 4.8. Diagramas de caja de los valores de $ecRSN$ para cada una de las clases

4.4.2. Selección de Características

Inicialmente, utilizando las 81 variables independientes disponibles, la red neuronal utilizada para seleccionar las características mostró un valor F_1 de 0,736 (F_{rsn} : 0,83, F_{fa} : 0,76, F_{or} : 0,61, F_{rd} : 0,48) una vez ajustados sus parámetros.

A continuación, la estrategia *FS* eligió un conjunto de 15 características, mostrando un rendimiento de F_1 de 0,746 (F_{rsn} : 0,85, F_{fa} : 0,79, F_{or} : 0,60, F_{rd} : 0,50), mejorando ligeramente el rendimiento de la red inicial.

La segunda estrategia de selección, *BE*, eliminó 63 características, obteniendo un conjunto final de 18 atributos igualando el rendimiento de la red inicial con un valor F_1 de 0,735 (F_{rsn} : 0,84, F_{fa} : 0,76, F_{or} : 0,59, F_{rd} : 0,50).

La Tabla 4.5 enumera los conjuntos de características seleccionadas por cada método (λ_{FS} y λ_{BE}), proveyendo además una abreviatura para cada una de ellas. Ahí se puede apreciar que las dos estrategias coincidieron en la selección de cinco características: (1) asimetría de la secuencia RR; (2 y 3) desviación media de RRd1 y *pNN50* máximo en el conjunto de ventanas deslizantes; (4) desviación estándar entre todas las distancias de los puntos del gráfico de Poincaré; (5) proporción entre el número de distintos patrones QRS encontrados y el número total de ondas R detectadas (PPC).

El número total de características seleccionadas teniendo en cuenta ambos métodos *FS* y *BE*, según las categorías establecidas es el siguiente: 4 estadísticas RR básicas utilizando toda la señal ECG; una característica basada en *pNNx* (*pNN50*); 17 características basadas en ventana deslizante; una característica basada en la gráfica de Poincaré; 2 basadas en la gráfica de Lorenz; una característica compleja (PPC); 2 características basadas en los vectores *pNNx* medios.

La Tabla 4.6 muestra la media y desviación estándar de los valores de las características seleccionadas según su clase. El comportamiento mostrado en prácticamente todas las variables es el mismo que el encontrado en aquellas basadas en las distancias al vector *pNNx* medio, donde existe separación en los rangos de valores de las clases RSN y FA, pero un solapamiento con las clases restantes. Estos solapamientos se producen principalmente entre la clase RSN y OR, por un lado, y entre las clases FA y RD por otro. No obstante, se seleccionaron 4 características que sí conseguían separar el ruido de las demás clases, principalmente de la FA. En concreto, estas características y sus valores de FA vs RD fueron los siguientes: (1) desviación media de RRd1 en el conjunto de ventanas deslizantes ($vR\sigma$ 255,1±150,5 vs 627,1±309,6); (2) desviación estándar entre todas las distancias de los puntos del gráfico de Poincaré ($Pcr\sigma$ 227,2±145 vs 682,5±403,4); (3) distancia mínima entre los puntos del gráfico de Lorenz ($Lrzmn$ 17,6±32,7 vs 178,7±357,1); (4) ratio de patrones QRS vs ondas R (PPC 0,14±0,15 vs 0,58±0,20). Destaca que, excepto la característica *Lrzmn*, las tres restantes fueron seleccionadas en ambos conjuntos λ_{FS} y λ_{BE} , mostrando su importancia en la detección de registros ruidosos.

Tabla 4.5. Características seleccionadas.Media (μ), desviación estándar (σ), curtosis (Ku), asimetría (Am).

Abrev.	Descripción	λ_{FS}	λ_{BE}
$RR\mu$	RR: μ .	X	
$RRAm$	RR: Am .	X	X
$RRd1\mu$	RRd1: μ .		X
$RRd1Ku$	RRd1: Ku .	X	
$pNN50$	pNN50		X
$vR\mu1$	Ventana deslizante – RR: μ media.		X
$vR\mu2$	Ventana deslizante – RR: μ máxima.	X	
$vR\sigma1$	Ventana deslizante – RR: σ media.	X	X
$vR\sigma2$	Ventana deslizante – RR: σ de la media.		X
$vRKu$	Ventana deslizante – RR: Ku media.		X
$vRd1\sigma1$	Ventana deslizante – RRd1: σ .		X
$vRd1\sigma2$	Ventana deslizante – RRd1: σ máxima		X
$vRd1Ku$	Ventana deslizante – RRd1: Ku mínima.	X	
$vRd2\sigma$	Ventana deslizante – RRd2: σ mínima.		X
$vSD\sigma$	Ventana deslizante – RMSSD: σ .	X	
$vSDmn$	Ventana deslizante – RMSSD mínimo.	X	
$vSDmx$	Ventana deslizante – RMSSD máximo		X
$vNN25\mu$	Ventana deslizante – $pNN25$ medio.	X	
$vNN75\mu$	Ventana deslizante – $pNN75$ medio.		X
$vNN25mx$	Ventana deslizante – $pNN25$ máximo.		X
$vNN50mx$	Ventana deslizante – $pNN50$ máximo.	X	X
$vNN75mx$	Ventana deslizante – $pNN75$ máximo.		X
$Pcr\sigma$	Poincaré – σ entre todas las distancias de los puntos	X	X
$Lrz\sigma$	Lorenz – σ distancias entre dos puntos consecutivos	X	
$Lrzmn$	Lorenz – Distancia mínima entre puntos consecutivos	X	
PPC	Complejas – Ratio de patrones QRS vs ondas R	X	X
$dtwRD$	Distancia DTW a la clase RD		X
$ecOR$	Error cuadrático respecto la clase OR	X	
Total		15	18

Tabla 4.6. Valores de las características seleccionadas según su clase ($\mu \pm \sigma$).

Abrev.	RSN	FA	OR	RD
<i>RRμ</i>	875,7 \pm 154,8	723,0 \pm 203,8	870,0 \pm 225,8	1090,1 \pm 338,6
<i>RRAm</i>	0,23 \pm 1,25	0,68 \pm 0,78	0,22 \pm 1,41	0,94 \pm 0,95
<i>RRd1μ</i>	-0,06 \pm 24,8	0,33 \pm 21,3	-0,06 \pm 22,5	5,05 \pm 124,4
<i>RRd1Ku</i>	4,34 \pm 3,55	3,23 \pm 1,63	5,21 \pm 4,12	3,53 \pm 1,63
<i>pNN50</i>	0,07 \pm 0,10	0,34 \pm 0,10	0,14 \pm 0,14	0,33 \pm 0,16
<i>vRμ1</i>	889,1 \pm 152,5	733,8 \pm 200,3	879,8 \pm 220,0	1056,6 \pm 243,7
<i>vRμ2</i>	966,2 \pm 224,6	815,6 \pm 253,8	970,9 \pm 280,2	1278,5 \pm 362,7
<i>vRσ1</i>	125,3 \pm 191,0	255,1\pm150,5	214,6 \pm 210,3	627,1\pm309,6
<i>vRσ2</i>	50,1 \pm 63,7	51,8 \pm 52,9	59,0 \pm 65,7	154,1 \pm 118,4
<i>vRKu</i>	2,53 \pm 0,74	2,98 \pm 0,98	2,91 \pm 1,00	2,65 \pm 1,07
<i>vRd1σ1</i>	114,2 \pm 176,5	98,6 \pm 110,1	146,5 \pm 155,9	355,8 \pm 286,1
<i>vRd1σ2</i>	316,1 \pm 465,4	407,8 \pm 310,7	449,9 \pm 430,3	1133,5 \pm 706,5
<i>vRd1Ku</i>	1,62 \pm 0,36	1,79 \pm 0,44	1,65 \pm 0,43	1,68 \pm 0,62
<i>vRd2σ</i>	40,0 \pm 97,3	224,1 \pm 149,7	111,1 \pm 193,5	297,2 \pm 299,8
<i>vSDσ</i>	103,3 \pm 156,0	90,5 \pm 98,1	132,7 \pm 136,5	307,4 \pm 242,9
<i>vSDmn</i>	31,73 \pm 72,7	139,9 \pm 93,5	73,8 \pm 122,4	287,9 \pm 186,6
<i>vSDmx</i>	287,3 \pm 412,0	379,5 \pm 277,1	410,5 \pm 377,6	1010,7 \pm 593,5
<i>vNN25μ</i>	0,16 \pm 0,14	0,41 \pm 0,10	0,20 \pm 0,15	0,36 \pm 0,17
<i>vNN75μ</i>	0,06 \pm 0,09	0,30 \pm 0,10	0,13 \pm 0,13	0,30 \pm 0,16
<i>vNN25mx</i>	0,30 \pm 0,21	0,54 \pm 0,11	0,35 \pm 0,21	0,51 \pm 0,24
<i>vNN50mx</i>	0,18 \pm 0,19	0,48 \pm 0,13	0,28 \pm 0,20	0,46 \pm 0,23
<i>vNN75mx</i>	0,14 \pm 0,17	0,43 \pm 0,13	0,26 \pm 0,20	0,44 \pm 0,22
<i>Pcrσ</i>	109,2 \pm 210,9	227,2\pm145,1	209,7 \pm 233,5	682,5\pm403,4
<i>Lrzσ</i>	136,2 \pm 282,0	280,5 \pm 188,9	255,1 \pm 293,4	854,9 \pm 537,1
<i>Lrzmn</i>	5,57 \pm 55,38	17,6\pm32,7	5,88 \pm 35,3	178,7\pm357,1
<i>PPC</i>	0,14 \pm 0,16	0,14\pm0,15	0,13 \pm 0,14	0,58\pm0,20
<i>dtwRD</i>	3,17 \pm 0,8	1,20 \pm 0,69	2,58 \pm 1,12	1,68 \pm 1,18
<i>ecOR</i>	2,00 \pm 1,98	5,59 \pm 4,43	2,69 \pm 3,66	9,41 \pm 13,41

4.4.3. Redes Neuronales

La Tabla 4.7 muestra los cinco mejores resultados alcanzados con cada conjunto de características seleccionadas mediante FFNN. Utilizando λ_{FS} , el valor F_1 más elevado fue de 0,751, usando 256 unidades en la capa oculta, un *learning rate* (α) de 0,6 y un *momentum* (β) de 0,5.

Por otro lado, utilizando el conjunto λ_{BE} , el valor F_1 más elevado fue de 0,749, alcanzado con una red neuronal de una capa oculta compuesta por 512 unidades, y valores de los parámetros α y β de 0,5 y 0,4 respectivamente.

Destaca que, con el conjunto λ_{FS} se obtuvieron los mejores resultados con una sensible diferencia respecto aquellos obtenidos con el conjunto λ_{BE} . Esta mejora sucedió utilizando menos características y número de unidades (256) que con el conjunto λ_{BE} , donde las redes con mejor rendimiento necesitaron entre 512 y 1024 unidades. Por lo tanto, utilizar un número menor de unidades en el mejor modelo λ_{FS} requirió menos carga computacional, tanto para el entrenamiento como para la clasificación de nuevas muestras.

Tabla 4.7. Mejores resultados obtenidos con redes neuronales durante la validación cruzada.

Número de unidades de la capa oculta ($\#u$), *learning rate* (α), *momentum* (β), conjunto de características (λ).

λ	$\#u$	α	β	F_1	F_{rsn}	F_{fa}	F_{or}	F_{rd}
λ_{FS}	256	0,6	0,5	0,751	0,854	0,759	0,640	0,304
	512	0,7	0,2	0,750	0,855	0,765	0,631	0,427
	256	0,5	0,4	0,750	0,857	0,757	0,635	0,373
	512	0,7	0,1	0,750	0,856	0,764	0,629	0,446
	256	0,5	0,3	0,750	0,857	0,758	0,635	0,356
λ_{BE}	512	0,5	0,4	0,749	0,859	0,750	0,639	0,434
	512	0,6	0,0	0,749	0,857	0,748	0,642	0,439
	512	0,8	0,1	0,749	0,857	0,749	0,641	0,414
	512	0,5	0,1	0,749	0,859	0,749	0,639	0,437
	1024	0,5	0,0	0,748	0,856	0,756	0,634	0,389

Respecto a los resultados de clasificación medios teniendo en cuenta el criterio global F_1 y el número de unidades de las FFNN, se obtuvieron valores de 0,733 con 64 unidades, 0,740 con 128 unidades, 0,744 con 256 unidades, 0,745 con 512 unidades, y 0,742 con 1024 unidades. La Figura 4.9 muestra los diagramas de cajas de estos resultados obtenidos durante la experimentación para el ajuste de parámetros, teniendo en cuenta los dos conjuntos de características λ_{FS} y λ_{BE} .

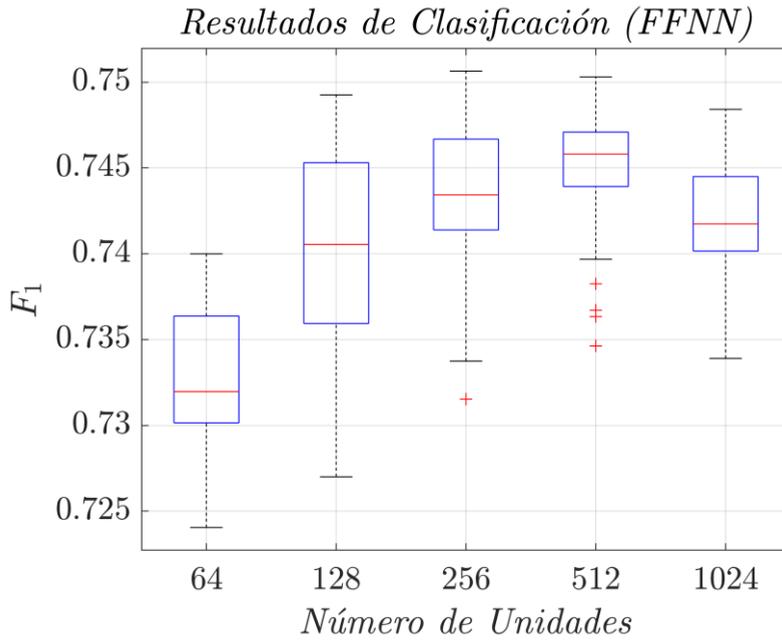


Figura 4.9. Diagramas de cajas de los resultados de clasificación con FFNN según el número de unidades utilizadas durante la experimentación.

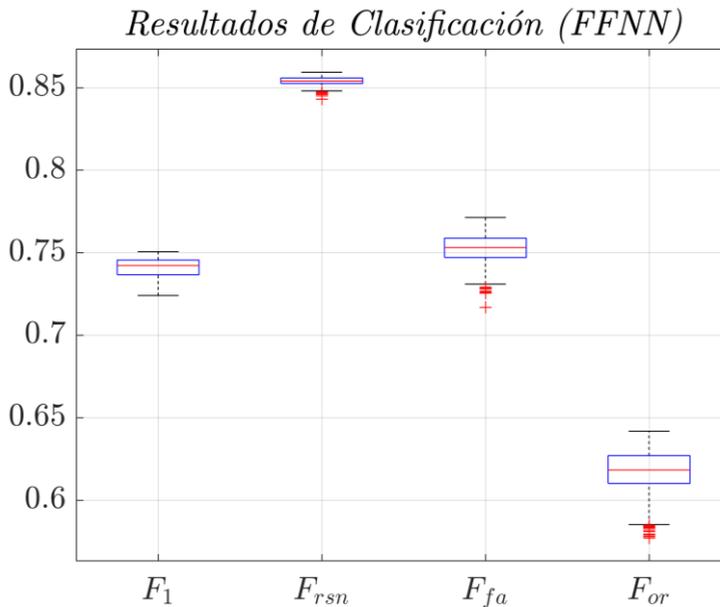


Figura 4.10. Diagramas de cajas de los resultados de clasificación con FFNN según el criterio global F_1 , y las medidas individuales F_{rsn} , F_{fa} y F_{or} .

Finalmente, la Figura 4.10 presenta en forma de diagramas de caja los resultados del criterio global F_1 , así como de las medidas individuales F_{rsn} , F_{fa} y F_{or} , obtenidos mediante los dos conjuntos de características λ_{FS} y λ_{BE} durante la experimentación para el ajuste de parámetros de las FFNN. Los valores medios de cada criterio de evaluación fueron de 0,74, 0,85, 0,75, 0,62, y de 0,38 para F_1 , F_{rsn} , F_{fa} , F_{or} y F_{rd} respectivamente. En todas estas medidas, la desviación estándar máxima fue de 0,02, lo que indica variaciones muy pequeñas entre todo el conjunto de resultados obtenidos.

En resumen, el grupo con mayor sensibilidad de detección utilizando FFNN fue el de ritmo sinusal normal, seguido cerca por el de FA. Por otro lado, la detección de otros ritmos anómalos fue considerablemente inferior a los anteriores, provocando un descenso en el rendimiento global de clasificación. Finalmente, la detección de registros ruidosos tuvo el menor rendimiento entre todas las clases.

4.4.4. Naïve Bayes

Los mejores resultados con modelos NB se consiguieron utilizando probabilidades empíricas para las distribuciones *a priori* de las clases, tanto en los modelos entrenados con el conjunto de datos λ_{FS} , como con λ_{BE} indistintamente. Además, esto sucedió usando siempre una función de densidad *KDF* (basada en *Kernel*) para modelar las distribuciones de probabilidad de cada una de las características.

La Tabla 4.8 muestra los cinco mejores resultados obtenidos con cada conjunto de características seleccionadas. Utilizando λ_{FS} , el valor F_1 más elevado fue de 0,672, usando valores de *bandwidth* (W) de 0,05 para FA (W_{fa}) y 0,02 para el resto de clases.

Por otro lado, utilizando el conjunto λ_{BE} , el valor F_1 más elevado fue de 0,66, usando valores de *bandwidth* de 0,005 en todas las clases. Se observó que no existía un patrón definido que relacionase el rendimiento de clasificación con los parámetros W de los modelos.

La Figura 4.11 presenta en forma de diagramas de caja los resultados del criterio global F_1 , así como de las medidas individuales F_{rsn} , F_{fa} y F_{or} , obtenidos mediante los dos conjuntos de características λ_{FS} y λ_{BE} durante la experimentación para el ajuste de parámetros en NB. Los valores medios de cada criterio de evaluación fueron de 0,62, 0,73, 0,64, 0,49, y de 0,31 para F_1 , F_{rsn} , F_{fa} , F_{or} y F_{rd} respectivamente. En todas estas medidas, la desviación estándar máxima fue de 0,06, lo que indica ligeras variaciones en todo el conjunto de resultados obtenidos.

En resumen, la sensibilidad de detección tuvo valores más elevados en el ritmo sinusal normal, y menores en FA, otros ritmos anómalos e identificando registros ruidosos al utilizar NB.

Tabla 4.8. Mejores resultados obtenidos con Naïve Bayes durante la validación cruzada.

Probabilidades de clase *a priori* calculadas empíricamente. Funciones de densidad basadas en Kernel: *bandwidth* para cada una de las clases (W). Conjunto de características (λ).

λ	$[W_{rsn}, W_{fa}, W_{or}, W_{rd}]$	F_1	F_{rsn}	F_{fa}	F_{or}	F_{rd}
λ_{FS}	[0,020, 0,050, 0,020, 0,020]	0,672	0,790	0,686	0,540	0,337
	[0,020, 0,050, 0,020, 0,035]	0,672	0,790	0,686	0,539	0,366
	[0,020, 0,050, 0,020, 0,050]	0,672	0,790	0,687	0,539	0,378
	[0,005, 0,050, 0,005, 0,020]	0,671	0,792	0,686	0,534	0,352
	[0,005, 0,050, 0,005, 0,050]	0,670	0,791	0,690	0,530	0,356
λ_{BE}	[0,005, 0,005, 0,005, 0,005]	0,660	0,775	0,658	0,545	0,258
	[0,005, 0,035, 0,005, 0,035]	0,658	0,774	0,665	0,534	0,355
	[0,005, 0,050, 0,005, 0,050]	0,657	0,773	0,664	0,535	0,349
	[0,005, 0,035, 0,005, 0,050]	0,657	0,773	0,665	0,533	0,353
	[0,005, 0,050, 0,005, 0,035]	0,657	0,777	0,663	0,535	0,353

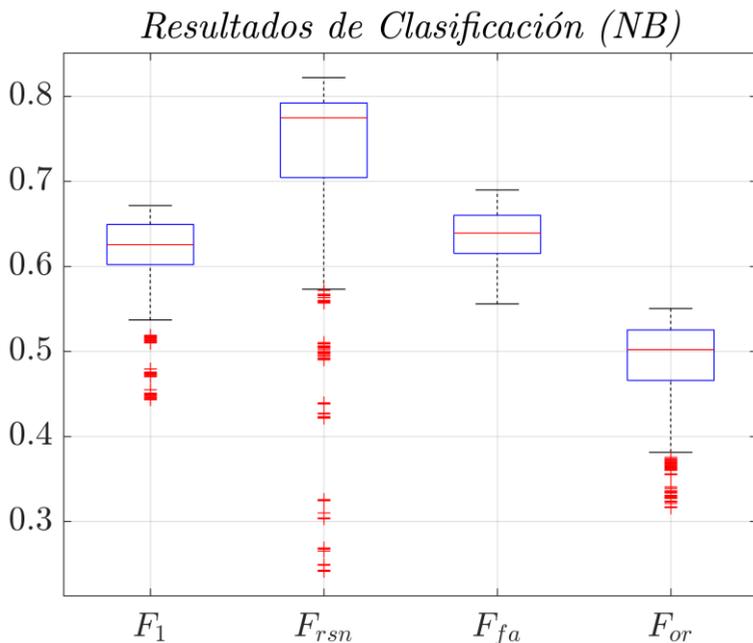


Figura 4.11. Diagramas de cajas de los resultados de clasificación con NB según el criterio global F_1 , y las medidas individuales F_{rsn} , F_{fa} y F_{or} .

4.4.5. Support Vector Machines

La Tabla 4.9 muestra los cinco mejores resultados alcanzados con cada conjunto de características seleccionadas mediante SVM. Utilizando λ_{FS} , el valor F_1 más elevado fue de 0,733, usando un valor de C de 1,0 y una función *kernel* de base radial gaussiana con un valor de γ de 1,44.

Por otro lado, utilizando el conjunto λ_{BE} , el valor F_1 más elevado fue de 0,725, usando un valor ajustado de C de 1,2 y una función *kernel* de base radial gaussiana con un valor de γ de 1,36.

La Figura 4.12 presenta en forma de diagramas de caja los resultados del criterio global F_1 , así como de las medidas individuales F_{rsn} , F_{fa} y F_{or} , obtenidos mediante los dos conjuntos de características λ_{FS} y λ_{BE} durante la experimentación para el ajuste de parámetros en modelos SVM. Los valores medios de cada criterio de evaluación fueron de 0,72, 0,82, 0,71, 0,61, y de 0,42 para F_1 , F_{rsn} , F_{fa} , F_{or} y F_{rd} respectivamente. En todas estas medidas, la desviación estándar máxima fue de 0,01, excepto para registros ruidosos donde fue de 0,07. Esto indica pocas variaciones en todo el conjunto de resultados, y por lo tanto en el espacio de búsqueda de parámetros.

Tabla 4.9. Mejores resultados obtenidos con SVM durante la validación cruzada.

Soft margin (C). Kernels: funciones gaussianas de base radial (RBF). Factor de escala en la función RBF (γ). Conjunto de características (λ).

λ	C	γ	F_1	F_{rsn}	F_{fa}	F_{or}	F_{rd}
λ_{FS}	1,0	1,44	0,733	0,835	0,736	0,629	0,435
	1,0	1,3	0,732	0,835	0,739	0,624	0,405
	1,0	1,4	0,732	0,836	0,734	0,627	0,425
	0,9	1,7	0,732	0,835	0,731	0,630	0,505
	1,0	1,7	0,732	0,834	0,732	0,629	0,482
λ_{BE}	1,2	1,36	0,725	0,835	0,726	0,614	0,337
	0,7	1,3	0,725	0,833	0,722	0,620	0,372
	1,0	1,5	0,725	0,835	0,721	0,618	0,386
	0,6	1,3	0,725	0,831	0,719	0,624	0,403
	0,7	1,4	0,724	0,831	0,721	0,621	0,424

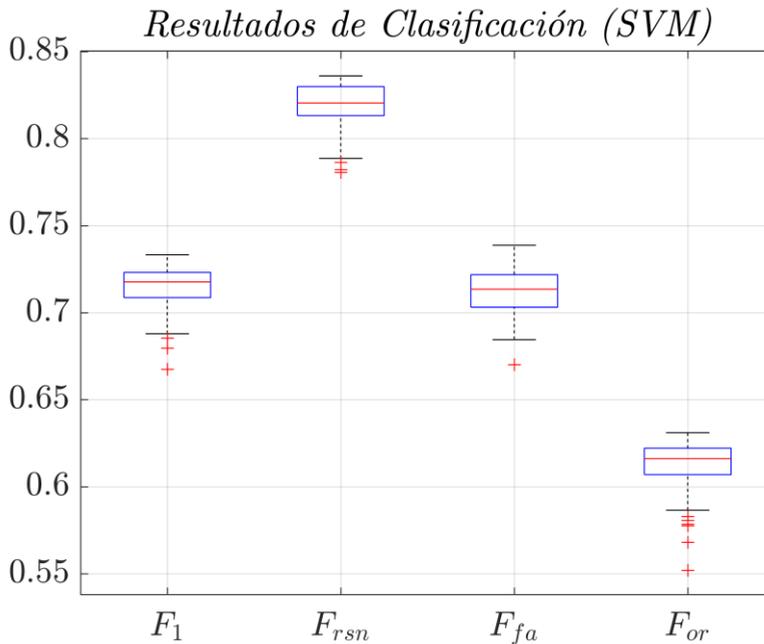


Figura 4.12. Diagramas de cajas de los resultados de clasificación con SVM según el criterio global F_1 , y las medidas individuales F_{rsn} , F_{fa} y F_{or} .

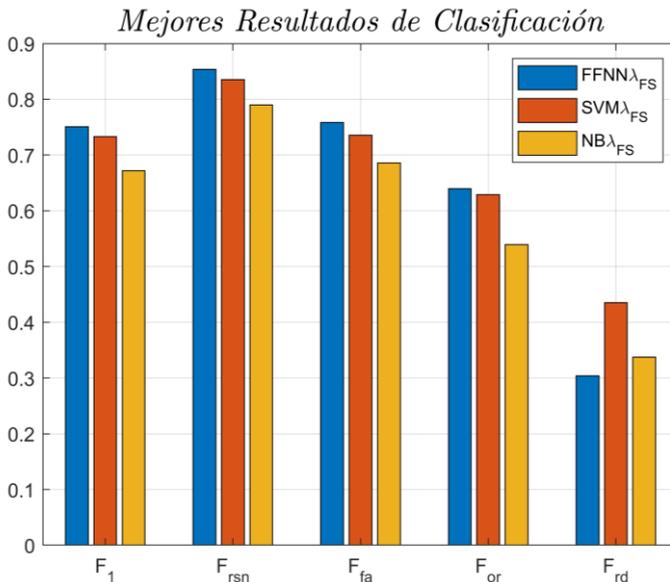
4.4.6. Visión global de los resultados

La Tabla 4.10 muestra el ranking respecto el criterio F_1 de los modelos con mejor rendimiento según el conjunto de características utilizado. En ella se observa que todos los modelos mostraron siempre un valor F_1 ligeramente más elevado al utilizar el conjunto de características λ_{FS} , el cual contenía 15 variables frente a las 18 de λ_{BE} .

Teniendo en cuenta únicamente el conjunto λ_{FS} , el primer puesto del ranking lo ocupó el modelo $FFNN\lambda_{FS}$ con un valor F_1 de 0,751, seguido de cerca por el modelo $SVM\lambda_{FS}$ con un valor F_1 de 0,733. El último puesto de este ranking lo ocupó el modelo $NB\lambda_{FS}$ con un valor F_1 de 0,672. Este ranking se repite también para las puntuaciones individuales F_{rsn} , F_{fa} y F_{or} . No obstante, en la detección de registros ruidosos, el modelo $SVM\lambda_{FS}$ obtuvo la mejor puntuación de toda la experimentación, con un valor F_{rd} de 0,435. La Figura 4.13 ilustra estas tendencias en los mejores resultados de clasificación según el tipo de modelo utilizado.

Tabla 4.10. Ranking de los mejores resultados obtenidos según el criterio F_1 .

Ranking	Modelo	F_1	F_{rsn}	F_{fa}	F_{or}	F_{rd}
1	$FFNN\lambda_{FS}$	0,751	0,854	0,759	0,640	0,304
2	$FFNN\lambda_{BE}$	0,749	0,859	0,750	0,639	0,434
3	$SVM\lambda_{FS}$	0,733	0,835	0,736	0,629	0,435
4	$SVM\lambda_{BE}$	0,725	0,835	0,726	0,614	0,337
5	$NB\lambda_{FS}$	0,672	0,790	0,686	0,540	0,337
6	$NB\lambda_{BE}$	0,660	0,775	0,658	0,545	0,258

**Figura 4.13. Resultados de los tres mejores modelos de clasificación obtenidos ($FFNN\lambda_{FS}$, $SVM\lambda_{FS}$, $NB\lambda_{FS}$), en cada uno de los criterios de evaluación.**

Finalmente, teniendo en cuenta el criterio global F_1 y el tipo de modelo utilizado, se obtuvieron valores medios de 0,741 con redes neuronales, 0,716 con SVM, 0,619 con NB. La Tabla 4.11 resume estos valores medios, así como las medidas de clasificación individuales para cada clase.

Así mismo, la Figura 4.14 muestra los diagramas de cajas de los valores F_1 obtenidos durante el ajuste de parámetros para cada tipo de modelo, teniendo en cuenta los dos conjuntos de características λ_{FS} y λ_{BE} . En esta figura se aprecia un ligero solapamiento entre las FFNN y SVM. Por otro lado, NB queda muy por debajo de los dos clasificadores anteriores, no existiendo solapamiento alguno en sus valores de F_1 .

Tabla 4.11. Resultados de clasificación obtenidos según el tipo de modelo (media±desviación estándar).

Modelo	F_1	F_{rsn}	F_{fa}	F_{or}	F_{rd}
<i>FFNN</i>	0,741±0,01	0,854±0,00	0,753±0,01	0,616±0,02	0,382±0,04
<i>SVM</i>	0,716±0,01	0,820±0,01	0,713±0,01	0,614±0,01	0,415±0,07
<i>NB</i>	0,619±0,04	0,728±0,12	0,637±0,03	0,491±0,05	0,306±0,06

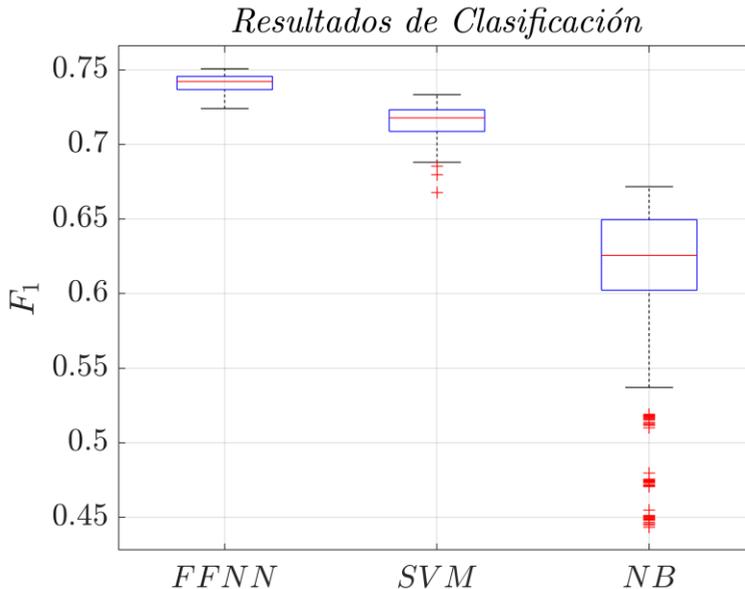


Figura 4.14. Diagramas de cajas de los resultados de clasificación obtenidos según el modelo utilizado durante la experimentación.

4.4.7. Costes Computacionales

La extracción de características tomó 116 segundos para las 8.496 muestras que no presentaron errores, lo que supuso un rendimiento de 73 muestras por segundo, o lo que es lo mismo, fueron necesarios 13,7 milisegundos de procesado por muestra.

Las Figura 4.15 y 4.16 ilustran los rendimientos computacionales de cada modelo en gráficos de barras. Estos costes temporales de clasificación, no tienen en cuenta la extracción de características.

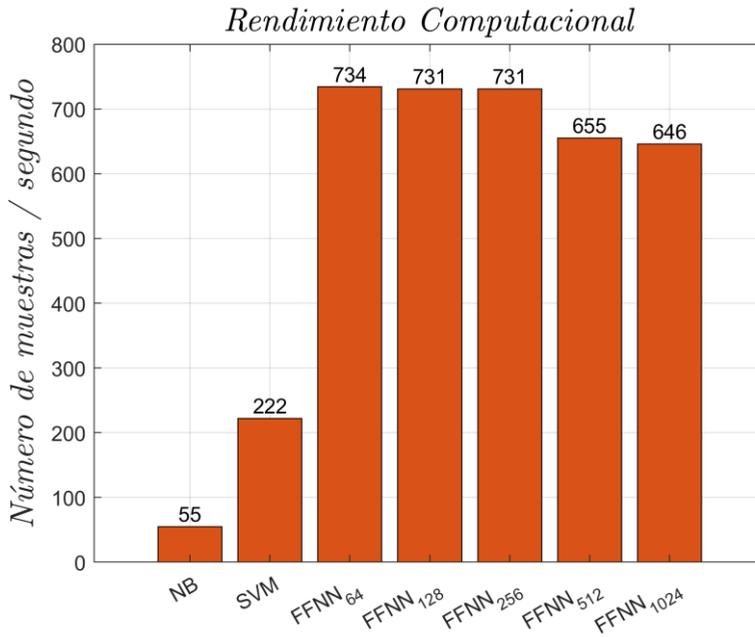


Figura 4.15. Rendimiento computacional medio de los modelos de clasificación (muestras/segundo).

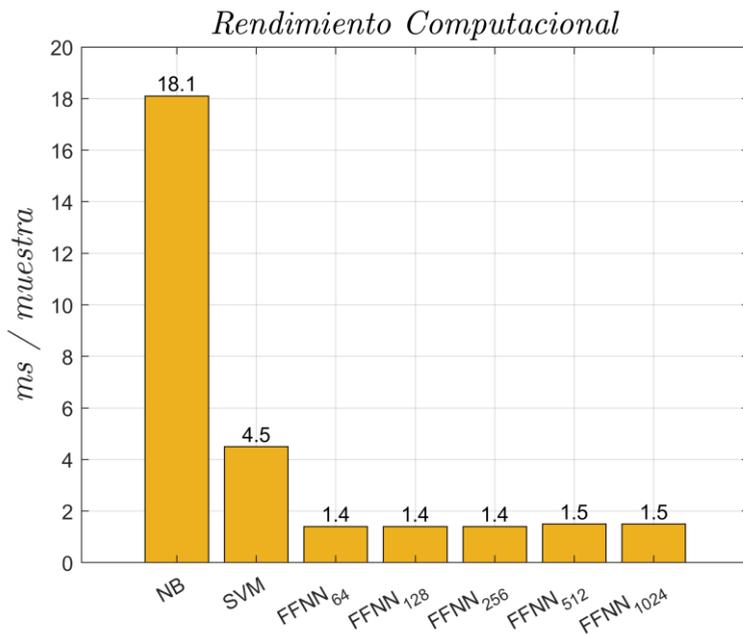


Figura 4.16. Rendimiento computacional medio de los modelos de clasificación (ms/muestra).

Encontramos que los modelos con mejor rendimiento computacional fueron las redes neuronales compuestas por 64, 128 y 256 unidades, pudiendo alcanzar rendimientos medios de 731 muestras por segundo (1,4 ms por muestra). Le siguieron de cerca las redes de 512 y 1024 unidades, con rendimientos computacionales alrededor de las 650 muestras por segundo (1,5 ms por muestra).

Más alejados de este rendimiento se encontraron los modelos SVM, procesando 222 muestras de media por cada segundo (4,5 ms por muestra).

Finalmente, los modelos NB procesaron 55 muestras por segundo (18,5 ms por muestra), teniendo el rendimiento computacional más bajo de todos los modelos con los que se experimentó.

De estos resultados se desprende que las redes neuronales aportaron una aceleración temporal media de 3,31 respecto los modelos SVM, y de 13,35 respecto NB.

En resumen, se necesitaron 15,1 ms de media para clasificar una muestra utilizando el modelo $FFNN\lambda_{FS}$, el cual presentó los mejores rendimientos de clasificación. Este valor se obtiene al sumar el coste computacional de la extracción de características, y el empleado por la red neuronal durante la clasificación.

4.5. Discusión

En este estudio se han propuesto un conjunto de indicadores y modelos de clasificación con el fin de detectar la FA en registros ECG de corta duración y una única derivación, etiquetando las muestras en cuatro grupos de ritmos cardiacos distintos: SNR, FA, OR y RD.

Los siguientes apartados, discuten los resultados obtenidos, comparándolos finalmente con trabajos previos.

4.5.1. Características basadas en los vectores $pNNx$ medios

Los resultados del análisis de las características basadas en distancias a los vectores $pNNx$ medios sugieren que para conseguir un nivel de separación óptimo entre registros de la clase RSN de aquellos de la clase FA, basta con usar alguna de las siguientes características propuestas: $dtwRSN$, $dtwFA$, $ecRSN$, $ecFA$. Además, se encontró que las características $pNN50$ o $pNN75$ también aportan un elevado nivel de separación entre RSN y FA, sin requerir del cálculo de distancias anterior. Sin embargo, es imposible distinguir utilizando únicamente este grupo de características, el RSN de otras patologías, y la FA de los registros ruidosos, al solaparse el rango de valores obtenidos mutuamente.

4.5.2. Selección de características

Los métodos de selección de características empleados consiguieron reducir en al menos un 78% el número de variables utilizadas manteniendo el rendimiento de clasificación. Respecto a las distribuciones de los valores de estas características seleccionadas, presentó un buen nivel de separación entre las clases RSN y FA, pero existieron solapamientos con las demás clases. No obstante, encontramos cuatro características que consiguieron aislar los registros ruidosos (clase RD) de las demás clases, principalmente de la FA que es con quien más se enmascaró generalmente. Éstas fueron las siguientes: (1) desviación media de $RRd1$ en el conjunto de ventanas deslizantes; (2) desviación estándar entre todas las distancias de los puntos del gráfico de Poincaré; (3) distancia mínima entre los puntos del gráfico de Lorenz; (4) ratio de patrones QRS frente al número de ondas R.

Destaca también que, en ambos grupos de variables seleccionadas (λ_{FS} y λ_{BE}), siempre resultó escogida alguna característica basada en las medidas de distancia a los vectores $pNNx$ medios ($dtwRD$ y $ecOR$), diseñadas durante este estudio. Esto sugiere que dichos criterios son útiles a la hora de discriminar parte de los registros ruidosos, así como otros ritmos patológicos.

4.5.3. Resultados de clasificación y costes computacionales

El rendimiento de los clasificadores entrenados con el conjunto de 15 variables λ_{FS} siempre resultó ligeramente superior que el conseguido con las 18 variables del conjunto λ_{BE} . Esta mejora se sitúa en un incremento máximo de 0,01 del valor de F_1 en los mejores resultados de modelos del mismo tipo. Aunque no es un incremento considerable, debe tenerse en cuenta que se produce utilizando tres características menos, reduciendo sensiblemente los costes computacionales de entrenamiento y clasificación.

El rendimiento de clasificación más elevado se obtuvo mediante el modelo $FFNN\lambda_{FS}$ ($F_1: 0,75$), seguido de cerca por el modelo $SVM\lambda_{FS}$ ($F_1: 0,73$). El modelo basado en Naïve Bayes $NB\lambda_{FS}$, no alcanzó un rendimiento de clasificación adecuado en comparación con los demás modelos ($F_1: 0,67$), a pesar de presentar un nivel de sensibilidad elevado para los ritmos normales ($F_{r_{sn}}: 0,79$).

Los resultados sugieren que la mayor dificultad de este problema se encuentra en la identificación de otros ritmos anómalos distintos a la FA, así como la separación de registros ruidosos de aquellos patológicos (FA incluida).

Respecto los costes computacionales, encontramos que la extracción de características necesitó de 13,7 ms por muestra. A se coste debe sumársele el cómputo del modelo para la clasificación de la muestra, el cual fue de 1,4 ms por muestra para el caso del modelo $FFNN\lambda_{FS}$ compuesto por 256 unidades. En este sentido, se observó un coste computacional muy inferior de las redes neuronales respecto SVM y NB, donde procesar una muestra costó de media 4,5 y 18,1 ms respectivamente. Esta mejora viene producida principalmente porque las funciones de activación ReLU de las unidades en las FFNN simplemente deben calcular un máximo entre 0 y un producto vectorial, mientras que en SVM y NB es necesario el cálculo de diversas funciones exponenciales.

En conclusión, las FFNN fueron los modelos que mejores rendimientos computacionales y de clasificación presentaron en este estudio, lo que las hacen candidatas idóneas para ser implementadas en dispositivos móviles de bajo consumo energético.

4.5.4. Comparación con otros trabajos

Las experimentaciones y resultados presentados en este estudio se basan en la participación previa del autor en el *PhysioNet/Computing in Cardiology Challenge 2017*, descrita en [Jiménez 2107a]. Allí, nuestro mejor valor de F_1 alcanzado durante la validación cruzada en el conjunto de entrenamiento público de 8.528 muestras, fue de 0,7644. No obstante, ese valor se alcanzó con un etiquetado de las muestras anterior (etiquetado V2) al más actual y utilizado en este estudio (etiquetado V3). Esto, unido a una experimentación ligeramente diferente, explicaría la ligera

diferencia de resultados que tenemos respecto este trabajo, donde el mejor valor que alcanzamos en el modelo $FFNN\lambda_{FS}$ fue de 0,75.

Asimismo, en <https://physionet.org/content/challenge-2017> y [Clifford 2017], se listan las puntuaciones de los 67 participantes de la competición. Estas puntuaciones se obtuvieron utilizando un conjunto de *test* oculto compuesto por 3.658 registros ECG no accesible al público hasta la fecha. La Tabla 4.12 compara los resultados de nuestro trabajo con aquellos que ocuparon las primeras y últimas posiciones del *ranking* de puntuación, y que finalmente publicaron su metodología.

Las mejores puntuaciones se obtuvieron con diferentes enfoques que variaron desde el aprendizaje automático tradicional hasta el aprendizaje profundo, no observándose un algoritmo superior para este problema de clasificación. El rendimiento de las técnicas de aprendizaje automático tradicionales utilizadas en la competición fue superior en un ~3% en comparación con las puntuaciones máximas obtenidas por los métodos de aprendizaje profundo. Es posible que el tamaño del conjunto de datos no fuera lo suficientemente grande como para determinar la mejor metodología de clasificación [Parvaneh 2019].

Los cuatro participantes con mayor puntuación del *ranking* obtuvieron valores F_1 de 0,83.

[Zabihi 2017] utilizó extracción y selección de características junto un clasificador *Random Forest*. alcanzando valores F_1 de $0,82\pm 0,03$ sobre el conjunto de datos de entrenamiento en su validación cruzada.

En [Hong 2017] se combinó la extracción de características clásicas y obtenidas mediante redes convolucionales a las que llamaron *Deep Neural Networks*, alimentando con ellas un ensamblado de clasificadores XGBoost [Chen 2016]. Con ello alcanzaron un valor F_1 de 0,86 sobre el conjunto de datos de entrenamiento en su validación cruzada.

[Datta 2017] utilizó una serie de clasificadores binarios en cascada, donde en una primera etapa, un registro se clasifica en una de las dos clases intermedias propuestas (SNR+OR o FA+RD), y en la segunda etapa se aplica otro clasificador binario que realiza el etiquetado final. Utilizando como clasificador binario AdaBoost alcanzaron un valor F_1 de 0,83 sobre el conjunto de datos de entrenamiento en su validación cruzada.

En [Teijeiro 2017] se combinaron dos clasificadores: uno de tipo XGBoost [Chen 2016] que evaluaba las características globales del ritmo de todo el registro ECG; el segundo clasificador fue una red neuronal recurrente alimentada con las características morfológicas individuales de cada latido detectado. Combinando estos dos clasificadores alcanzaron un valores F_1 de 0,85 sobre el conjunto de datos de entrenamiento en su validación cruzada.

En nuestro caso, con [Jiménez 2017a] ocupamos la posición 34 del *ranking*, *i.e.* la mitad de la tabla de puntuaciones. El valor de F_1 que obtuvimos en el conjunto de *test* oculto fue de 0,77, lo que sugiere que nuestra metodología aplicada a FFNN no sobreajusta el rendimiento de clasificación al conjunto de entrenamiento.

Tabla 4.12. Comparación de nuestros resultados de clasificación con otros trabajos, utilizando el conjunto oficial de *test* oculto.

Estas puntuaciones fueron proporcionadas *online* en el contexto del *PhysioNet/Computing in Cardiology Challenge 2017* [Clifford 2017].

Referencia	(Ranking) Modelo Utilizado	F_1	F_{rsn}	F_{fa}	F_{or}	F_{rd}
Zabihi 2017	(1) Random Forest	0,83	0,91	0,84	0,73	0,50
Hong 2017	(1) Deep Neural Network	0,83	0,91	0,81	0,75	0,57
Datta 2017	(1) Cascaded Binary Classif.	0,83	0,92	0,82	0,75	0,52
Teijeiro 2017	(1) Abductive Interpretation	0,83	0,90	0,85	0,74	0,56
Bin 2017	(5) Decision Tree Ensemble	0,82	0,91	0,82	0,73	0,57
Zihlmann 2017	(5) CRNN	0,82	0,91	0,82	0,73	0,57
Xiong 2017	(5) Neural Networks	0,82	0,90	0,82	0,73	0,53
Patidar 2017	(9) Fourier-Bessel expansion	0,81	0,91	0,80	0,72	0,51
Jiménez 2017a	(34) FFNN	0,77	0,89	0,76	0,67	0,42
Hasna 2017	(62) Decision Tree	0,53	0,82	0,34	0,43	0,32

Teniendo en cuenta los 67 participantes de la competición, y las puntuaciones obtenidas en el conjunto de *test* oculto de 3.658 registros, los valores medios del criterio de evaluación global F_1 fueron de $0,73 \pm 0,12$. Por otro lado, los valores medios de los criterios de individuales por cada clase fueron de $0,87 \pm 0,07$ (F_{rsn}), $0,70 \pm 0,15$ (F_{fa}), $0,61 \pm 0,15$ (F_{or}), y de $0,43 \pm 0,19$ (F_{rd}) respectivamente. Esto indica que la tendencia general fue la misma para todos los participantes de la competición, teniendo los mayores problemas de clasificación en registros ruidosos y otros ritmos patológicos.

La Figura 4.17 muestra los diagramas de cajas de estos valores, apreciándose la tendencia descrita anteriormente, donde la sensibilidad de detección de las distintas clases, ordenada de mayor a menor, es la siguiente: F_{rsn} , F_{fa} , F_{or} , F_{rd} .

Nuestros resultados presentaron la misma tendencia, situándose por encima de la media en F_1 (+0,04), F_{rsn} (+0,02), F_{fa} (+0,06), F_{or} (+0,06) y F_{rd} (+0,01), respecto de los obtenidos por todos los competidores, lo que indica la validez de nuestra propuesta.

Además de los resultados de 2017, otros autores han utilizado el mismo conjunto de datos de entrenamiento y métricas de evaluación posteriormente para evaluar distintos modelos de clasificación.

[Nguyen 2021] alcanzó valores F_1 de 0,84 apilando clasificadores SVM alimentados con características estadísticas de segmentos de la señal producidas por una red neuronal convolucional.

[Pandey 2022] alcanzó valores F_1 de 0,8 combinando técnicas tradicionales de aprendizaje automático y redes neuronales profundas. Dicho modelo híbrido combinó una red neuronal convolucional (Red Residual (ResNet)), una red de memoria a corto plazo bidireccional a largo plazo (BLSTM) y una red neuronal de función de base radial (RBF).

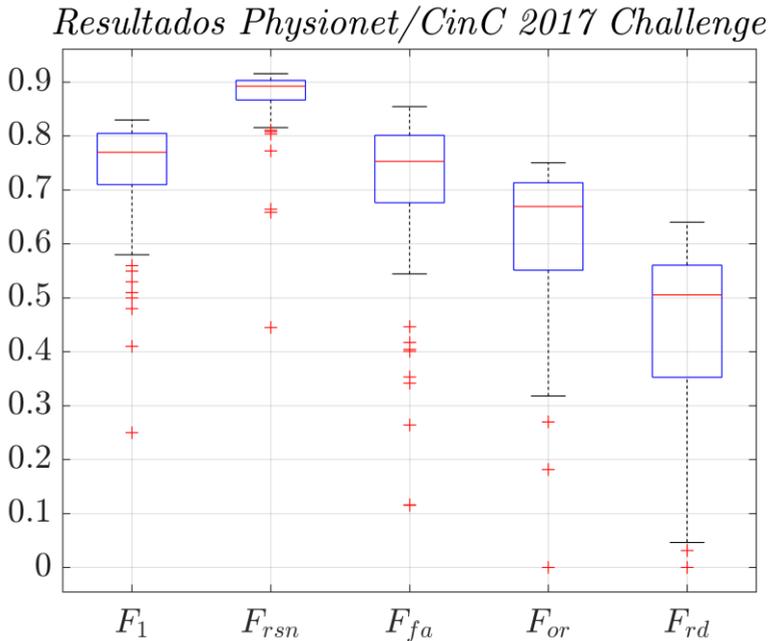


Figura 4.17. Diagramas de caja de los resultados de clasificación de los 67 participantes en el *PhysioNet/Computing in Cardiology Challenge 2017*.

Estos trabajos más recientes no presentaron grandes aumentos en los rendimientos de clasificación respecto los iniciales en el mismo conjunto de datos. No obstante, en [Rohr 2022] sí que se observa un aumento considerable de la puntuación F_1 obtenida en el conjunto de *test* oculto con un valor de 0,91. Para ello utilizaron un modelo al que llamaron ECG-RCLSTM-Net, el cual empleaba una red neuronal convolucional, una red residual (ResNet) y una memoria a corto plazo (LSTM). Este último resultado se obtuvo, en parte, gracias a añadir 4.652 muestras adicionales procedentes de bases de datos externas para la validación de los modelos de clasificación.

Destacamos aquí las conclusiones de la revisión de [Parvaneh 2019], donde se apuntó que se requieren esfuerzos adicionales por la comunidad científica para mejorar la comparación entre diferentes algoritmos (tanto en rendimiento como en eficiencia computacional), evaluar la generalización de los algoritmos propuestos y mejorar la interpretabilidad de los modelos basados en aprendizaje profundo.

En este sentido, ninguno de los estudios mencionados evaluó el rendimiento computacional, como hicimos nosotros. Los modelos de clasificación basados en redes convolucionales necesitan de un número elevado de capas y unidades, y los métodos de clasificación clásicos requieren también de extracción de características y clasificadores alimentados con ellas. Esto sugiere que nuestro coste computacional

utilizando FFNN durante la clasificación es menor o igual que cualquiera de estos trabajos.

4.5.5. Limitaciones y trabajos futuros

Este estudio presenta algunas limitaciones a tener en cuenta. A continuación, enumeramos algunas de ellas, sugiriendo ciertas líneas de trabajo futuras al respecto.

La principal limitación de este estudio es que, a pesar de presentar un buen nivel de separación entre el ritmo sinusal normal y la FA, es preciso aumentar la sensibilidad en la detección de ritmos anómalos distintos a la FA para conseguir un rendimiento global del sistema óptimo. Para ello se proponen las siguientes estrategias.

En primer lugar, es necesario abordar la extracción de otro tipo de características del ECG no relacionadas con la variabilidad del RR, como, por ejemplo, la amplitud y morfología del complejo QRS-T o de la onda P. También sería útil la inclusión de un detector de latidos ectópicos, y de ausencias y/o inversiones de las ondas T y P. Finalmente, sería interesante explorar la extracción de características relacionadas con el tiempo-frecuencia de la señal, teniendo en cuenta el incremento del coste computacional.

En segundo lugar, trabajos futuros podrían abordar la selección de características mediante otras técnicas, como los algoritmos genéticos, o el enfriamiento simulado, entre otras. Es probable que así se encontrasen nuevos conjuntos de variables independientes que mejorasen, todavía más, los rendimientos de clasificación obtenidos con los métodos de selección voraces utilizados en este estudio.

En tercer lugar, la detección de otros ritmos anómalos podría mejorarse mediante la ampliación del conjunto de entrenamiento para conseguir una mayor generalización en la clasificación. Un recurso a utilizar podría ser *PhysioBank Databases* de *PhysioNet* (<https://physionet.org/physiobank/database>) donde pueden encontrarse bases de datos con registros ECG ya etiquetados.

En último lugar, las ampliaciones futuras deberían abordar la inclusión de otros tipos de modelos de clasificación, principalmente basados en redes convolucionales, o modelos ocultos de Markov. Ambos tipos de modelos ya son utilizados con éxito en distintas aplicaciones médicas, bioinformáticas, o de reconocimiento del habla, entre otras.

4.5.6. Conclusiones

En este estudio se ha presentado y evaluado una metodología de clasificación de registros ECG, de corta duración y una única derivación, en cuatro clases de ritmos cardiacos distintos: ritmo sinusal normal, fibrilación auricular, otros ritmos anómalos, y registros ruidosos.

Para ello combinamos la extracción de características del ECG basadas en la variabilidad del intervalo RR, la selección de conjuntos óptimos de características, y el entrenamiento y validación de tres tipos de modelos de clasificación: FFNN, SVM y NB.

Después de un análisis sistemático, los resultados mostraron que los modelos con mejor rendimiento de clasificación y computacional fueron las FFNN (F_1 : 0,75; 1,4 ms/muestra), seguidas de cerca por los modelos SVM (F_1 : 0,73; 4,5 ms/muestra). Por otro lado, NB no consiguió rendimientos adecuados en comparación con los modelos anteriores (F_1 : 0,67; 18,1 ms/muestra).

Los resultados también mostraron un nivel óptimo de separación entre RSN y FA, pero poca sensibilidad en la identificación de otros ritmos anómalos distintos a la FA, así como de los registros ruidosos, lo que hizo bajar el rendimiento global de clasificación del sistema. Mejorar este aspecto incorporando características más específicas para otras cardiopatías es una de las direcciones a explorar en ampliaciones futuras de este trabajo.

En resumen, los modelos desarrollados presentan un bajo consumo computacional durante la clasificación, y un buen rendimiento en la identificación de FA. Por lo tanto, podrían ser candidatos adecuados para ser incluidos en sistemas que utilicen dispositivos portátiles para la monitorización cardiaca, o de gestión de pacientes, con el objetivo de ayudar en la detección temprana de arritmias auriculares en cribados iniciales.

Capítulo 5

DetECCIÓN DE MÚLTIPLES AFECCIONES CARDIACAS EN DIVERSOS CONJUNTOS DE DERIVACIONES

5.1. Introducción

La relevancia clínica de las enfermedades cardíacas, estrechamente vinculadas con el envejecimiento de la población, está aumentando junto con su incidencia y prevalencia [Savarese 2022, Chow 2012], y como consecuencia se está convirtiendo en la principal causa de muerte mundial [Vaduganathan 2022, Virani 2021, Gaziano 2010]. Además de disminuir la calidad de vida de los pacientes, todas estas patologías provocan un coste muy elevado a los sistemas nacionales de salud [Kumar 2022, Birger 2021, Gheorghe 2018, Wang 2015, Montes 2013].

En este sentido, el ECG de 12 derivaciones es la técnica principal en el diagnóstico cardíaco [Kligfield 2002], siendo una herramienta cribado y detección de patologías barata y ampliamente disponible. Sin embargo, este proceso requiere de médicos experimentados para interpretar estos registros de ECG, siendo una tarea que consume tiempo, recursos y que además está sujeta a la variabilidad entre observadores [Bickerton 2019].

Por otro lado, los dispositivos inteligentes con conjuntos de derivaciones reducidos pueden ser útiles como herramientas de detección masiva y de reducción de costes hospitalarios [Hermans 2022, Duncker 2021]. El principal desafío en el diagnóstico con estos dispositivos es encontrar un equilibrio adecuado entre la sensibilidad (para no pasar por alto problemas cardíacos importantes) y la especificidad (para evitar colapsar innecesariamente los centros de salud).

En consecuencia, es de especial relevancia diseñar e implementar algoritmos precisos adecuados para dispositivos móviles o portátiles capaces de detectar, más allá de únicamente la FA, múltiples afecciones cardíacas en registros de ECG.

5.1.1. Objetivos

En este estudio, nuestro objetivo principal es desarrollar y validar una metodología robusta y de bajo coste computacional, capaz de identificar una amplia variedad de afecciones cardíacas a partir de cinco combinaciones de derivaciones de ECG diferentes (12, 6, 4, 3 y 2 derivaciones) además del escenario de la derivación única I. Este objetivo intenta abordar la necesidad clínica de poder identificar automáticamente múltiples afecciones cardíacas usando nuevas tecnologías portátiles que trabajan con información electrocardiográfica limitada. También pretendemos evaluar el rendimiento de clasificación de nuestra solución con segmentos cortos de ECG, ya que son los más utilizados en dispositivos móviles e inteligentes.

Los objetivos detallados de este estudio son los siguientes:

- Definir y extraer automáticamente características espectrales y derivadas de la actividad ventricular de las señales ECG disponibles.
- Definir, validar y analizar procesos de selección de características.
- Entrenar y validar el rendimiento de clasificación de diversos modelos de clasificación binarios que utilice conjuntos de distintas derivaciones ECG, para identificar individualmente cada afección cardíaca en el conjunto de datos, seleccionando aquel modelo con mejor rendimiento en cada una de ellas.
- Definir y validar el rendimiento de clasificación global de una metodología robusta que utilice los clasificadores binarios anteriores, para etiquetar nuevas muestras entre 26 ritmos cardíacos diferentes, patológicos principalmente.

5.2. Materiales

Para validar el rendimiento del enfoque propuesto, nuestros experimentos se realizaron y examinaron en el contexto del *PhysioNet/Computing in Cardiology Challenge 2021* [Pérez 2020, Reyna 2021], donde se utilizaron siete bases de datos con un total de 131.149 muestras de ECG de 12 derivaciones. De estas muestras, usamos los 88.253 registros de ECG proporcionados como conjunto de entrenamiento público, los cuales contenían también la edad y el sexo de cada paciente. Por otro lado, los organizadores de la competición evaluaron *online* el rendimiento de los modelos de clasificación con las bases de datos ocultas de validación y *test*, compuestas por 6.630 y 36.266 registros de ECG, respectivamente. Los datos crudos tienen siete orígenes diferentes mostrados en la Tabla 5.1. El primer

origen es la base de datos del *China Physiological Signal Challenge 2018* (CPSC) [Liu 2018]. El segundo origen es la base de datos del *St. Petersburg Institute of Cardiological Technics* (INCART) [Tihonenko 2008]. El tercer origen es la base de datos *Physikalisch-Technische Bundesanstalt database* (PTB) [Bousseljot 1995, Wagner 2020]. El cuarto origen es la base de datos del *Georgia 12-lead ECG Challenge database* (G12EC). La quinta base de datos proviene de una institución estadounidense no revelada (*Undisclosed*). El sexto origen son las bases de datos procedentes del *Chapman University Shaoxing People's Hospital* (Chapman-Shaoxing) [Zheng 2020b] y del *Ningbo First Hospital* (Ningbo) [Zheng 2020c] (CHAP-SHX). Finalmente, el séptimo origen es la base de datos UMich de la Universidad de Michigan.

Los registros públicos de ECG utilizados para el entrenamiento de modelos de clasificación en este trabajo provienen de cinco de las bases de datos mencionadas: CPSC, INCART, PTB, G12EC y CHAP-SHX. Se pueden encontrar explicaciones más detalladas de los registros (frecuencia de muestreo, duración de las muestras, etc.) en [Pérez 2020, Reyna 2021].

La Tabla 5.2 muestra las afecciones cardíacas a detectar y el número de muestras marcadas con dichas etiquetas en el conjunto de datos de entrenamiento. Aquí, un registro de ECG puede aparecer etiquetado con más de una clase; además algunas clases contemplan dos enfermedades cardíacas al puntuar como el mismo diagnóstico.

Tabla 5.1. Número de registros ECG de la base de datos utilizada, procedente del *PhysioNet/Computer in Cardiology Challenge (2021)*.

Base de datos	Entrenamiento (Pública)	Validación (Oculto)	Test (Oculto)	Total
CPSC	10.330	1.463	1.463	13.256
INCART	74	-	-	74
PTB	22.353	-	-	22.353
G12EC	10.344	5.167	5.161	20.672
<i>Undisclosed</i>	-	-	10.000	10.000
CHAP-SHX	45.152	-	-	45.152
UMich	-	-	19.642	19.642
Total	88.253	6.630	36.266	131.149

Tabla 5.2. Ritmos cardiacos a detectar (clases) en el conjunto de datos público para entrenamiento.

	Abreviatura	Nombre	Muestras para entrenamiento
1	AS	Arritmia Sinusal	3.790
2	BAVPG	Bloqueo Auriculoventricular de Primer Grado	3.534
3	BFAnI	Bloqueo Fascicular Anterior Izquierdo	2.186
4	BR	Bloqueo de Rama	522
5	Bradi	Bradicardia	295
6	BradiS	Bradicardia Sinusal	18.918
7	BRDC BRD	Bloqueo de Rama Derecha Completo Bloqueo de Rama Derecha	1.779 3.051
8	BRDI	Bloqueo de Rama Derecha Incompleto	1.857
9	BRIC BRI	Bloqueo de Rama Izquierda Completo Bloqueo de Rama Izquierda	213 1.281
10	CAP LSP	Contracciones Auriculares Prematuras Latidos Supraventriculares Prematuros	3.041 224
11	CVP LVP	Contracciones Ventriculares Prematuras Latidos Ventriculares Prematuros	1.279 659
12	DEjD	Desviación del Eje a la Derecha	1.280
13	DEjI	Desviación del Eje a la Izquierda	7.631
14	FA	Fibrilación Auricular	5.255
15	FLA	<i>Flutter</i> Auricular	8.374
16	MPrR	Mala Progresión de la Onda R	638
17	PRPr	Intervalo PR Prolongado	392
18	QAn	Onda Q Anormal	2.076
19	QTPr	Intervalo QT Prolongado	1.907
20	RMp	Ritmo de Marcapasos	1.481
21	RSN	Ritmo Sinusal Normal	28.971
22	TacS	Taquicardia Sinusal	9.657
23	TAn	Onda T anormal	11.716
24	TCII	Trastorno de Conducción Intraventricular Inespecífico	1.768
25	TInv	Onda T invertida	3.989
26	VQRSB	Voltaje QRS Bajos	1.599

Al compartir algunas enfermedades el ritmo y sus características principales, además de puntuar como el mismo diagnóstico en el *PhysioNet/Computer in Cardiology Challenge 2021* [Reyna 2021], volvimos a etiquetar las siguientes cuatro clases en el conjunto de datos: BRD como BRDC, BRI como BRIC, LSP como CAP y finalmente LVP como CVP. De esta manera, redujimos el número de clases a detectar de 30 a 26.

Por otro lado, se eliminaron del conjunto de datos de entrenamiento 6.287 registros de ECG que no pertenecían a ninguna de las 26 clases a detectar, ya que ningún clasificador binario de este estudio podría tomarlos como muestras positivas.

Finalmente, la Tabla 5.3 muestra el número y los conjuntos de derivaciones de ECG evaluados en este estudio, utilizados para entrenar y validar los modelos de clasificación. Además de los conjuntos de derivaciones oficiales del *PhysioNet/Computing in Cardiology Challenge 2021* (12, 6, 4, 3 y 2 derivaciones), también reportamos los resultados obtenidos utilizando únicamente la derivación I. Consideramos que esto es importante para evaluar la idoneidad de nuestros métodos de clasificación automática para señales de ECG en dispositivos que registran sólo la derivación I.

Tabla 5.3. Número y conjuntos de derivaciones ECG utilizadas para entrenar y validar nuestros modelos de clasificación.

Número de derivaciones	Conjunto de derivaciones
12	I, II, III, aVR, aVL, aVF, V1, V2, V3, V4, V5, V6
6	I, II, III, aVR, aVL, aVF
4	I, II, III, V2
3	I, II, V2
2	I, II
1	I

5.3. Métodos

En esta sección, describimos la metodología propuesta para identificar 25 anomalías del ECG además del Ritmo Sinusal Normal (RSN), sumando un total de 26 clases diferentes a detectar, usando las combinaciones de derivaciones de ECG que aparecen en la Tabla 3.

Primero presentamos el esquema de validación de los modelos y las reglas de puntuación aplicadas. A continuación, presentamos la reducción de ruido, el procesamiento y la extracción de características de las señales. Posteriormente, se describe la selección de características para la reducción de dimensionalidad en el conjunto de datos. Seguidamente, presentamos los procesos de entrenamiento y validación para las distintas configuraciones de los modelos de clasificación binaria propuestos. Para concluir, describimos el enfoque de clasificación *One-vs-Rest* utilizado con el objetivo de proporcionar un sistema de clasificación de múltiples clases destinado a detectar las 26 categorías distintas de afecciones cardiacas mencionadas anteriormente.

El entorno de trabajo donde realizamos todos estos pasos fue MATLAB (R2021a, The MathWorks); el código desarrollado, utilizado para entrenar y validar los modelos está disponible en https://github.com/sjimenezupv/itaca_upv.cinc2021.special_issue. Por otro lado, el hardware utilizado durante este estudio fue un ordenador con un procesador Intel(R) Core(TM) i9-10900K CPU @ 3.70GHz, 64 GB de memoria RAM, un disco SSD con 1TB de memoria disponible y una tarjeta gráfica NVIDIA GeForce RTX 3080 compatible con CUDA. Este ordenador tenía instalado el sistema operativo Windows 10 Pro (64 bits). En este trabajo, presentaremos nuestros costes computacionales en términos de tiempo utilizado durante los tres procesos principales (extracción de características, entrenamiento y validación de los modelos de clasificación) junto con el rendimiento de nuestro sistema en términos de muestras procesadas por segundo.

5.3.1. Esquema de Validación

Usamos una validación cruzada triple (de 3 conjuntos o pliegues) con las 88.253 muestras de entrenamiento disponibles, ya que la cantidad de muestras en la base de datos era suficientemente grande: para un número mayor de pliegues el tiempo de entrenamiento podía volverse innecesariamente elevado. Además, se seleccionó equitativamente entre las distintas bases de datos disponibles las muestras para cada pliegue, de manera que tuviéramos el mismo número de muestras procedentes de cada base de datos en cada pliegue. De este modo, para cada conjunto de la validación cruzada, utilizamos 58.835 (66,6%) muestras para entrenamiento y 29.418 (33,3%) para pruebas con el fin de evaluar el rendimiento de los modelos de clasificación, utilizando las combinaciones de 12, 6, 4, 3, 2 y 1 derivación ECG.

Estos rendimientos, serán identificados como resultados *offline*, ya que son obtenidos por nosotros mismos en nuestro propio entorno de trabajo, con los datos públicos de la competición.

Por otro lado, en este trabajo detallamos los resultados en los conjuntos de datos oficiales de validación y *test*, también denominados conjuntos de datos de validación y *test online*, ya que fueron resultados obtenidos por los organizadores de la competición en sus propios servidores. Estos conjuntos de datos contienen 6.630 registros para validación y 36.266 registros de *test* con 12 derivaciones ECG respectivamente. Estos resultados *online* fueron proporcionados por los organizadores utilizando las combinaciones oficiales de 12, 6, 4, 3 y 2 derivaciones. Los conjuntos de datos *online* no deben confundirse con el conjunto de datos de entrenamiento utilizado durante la validación cruzada *offline*. Cabe destacar que los conjuntos de datos oficiales de validación y *test* no estuvieron disponibles para entrenar ni validar ningún modelo de clasificación en este trabajo y se usaron únicamente para obtener las métricas de puntuación de nuestros modelos descritas a continuación. Nótese también que estos resultados *online* no incluyen ningún resultado de derivación única como el nuestro en la validación cruzada *offline*.

5.3.2. Reglas de Puntuación

En este estudio reportamos los resultados obtenidos con la regla de puntuación de la competición *PhysioNet/Computing in Cardiology Challenge 2021* descrita en [Reyna 2021]. Esta regla de puntuación, denominada *Challenge Metric (CM)*, utiliza una colección $C = [c_j]$ como una lista de diagnósticos positivos o negativos, calculando la matriz de confusión de varias clases $A = [a_{ij}]$, donde a_{ij} es el número de registros que fueron clasificados como pertenecientes a la clase c_i pero que realmente pertenecen a la clase c_j . Seguidamente, la matriz $W = [w_{ij}]$ pondera la matriz de confusión A , basándose en la similitud de tratamientos o diferencias en los riesgos entre distintas patologías. De esta forma, se obtiene la puntuación no normalizada, denominada S , usando la siguiente expresión

$$S = \sum_{ij} w_{ij} \times a_{ij} \quad (5.1)$$

A continuación, la puntuación S es normalizada para dar un valor de 1 en el caso de un clasificador que siempre identifica la clase o clases positivas correctamente, mientras que proporcionará un valor de 0 para un clasificador inactivo que siempre genera predicciones negativas para todas las clases existentes. Esta normalización da como resultado el valor final de la puntuación *CM*, obtenido mediante la siguiente proporción

$$CM = \frac{S - S_{inactive}}{S_{true} - S_{inactive}} \quad (5.2)$$

donde $S_{inactive}$ es la puntuación para un clasificador inactivo y S_{true} es la puntuación de un clasificador perfecto. Nótese que esta puntuación CM puede dar como resultado valores negativos si la proporción de falsos positivos es elevada, mostrando así una puntuación más baja que la de un clasificador que devuelve únicamente etiquetas negativas.

Por otro lado, también reportamos la media geométrica entre la sensibilidad y la especificidad tanto para cada clasificador binario individual como la media \pm la desviación estándar global para cada combinación de derivaciones. Esta puntuación, denominada aquí métrica G , se utilizó para seleccionar los clasificadores binarios con el mejor rendimiento durante el entrenamiento y validación en nuestra experimentación, correspondiéndose con la siguiente expresión

$$G = \sqrt{\text{Sensibilidad} \times \text{Especificidad}} \quad (5.3)$$

También reportamos el área bajo la curva ROC (AUROC) a pesar de que no usamos sus valores para validar ni seleccionar ningún hiperparámetro o modelo binario durante la etapa de entrenamiento.

5.3.3. Procesado de Señal

La duración de las señales de la base de datos es heterogénea, con una duración media de 10 a 1.800 segundos, mientras que su frecuencia de muestreo mayoritariamente es de 500 Hz [Pérez 2020, Reyna 2021, Reyna 2022, Zheng 2020b, Zheng 2020c]. El preprocesamiento de la señal que se describe a continuación se aplicó a todas los registros ECG independientemente de su duración.

En primer lugar, todas las señales ECG se remuestrearon a 500 Hz si esa no era su frecuencia de muestreo original utilizando la función *resample* de Matlab R2021a, homogeneizando así al valor más frecuente en la base de datos.

A continuación, aplicamos un filtro pasa banda de *Butterworth* entre 0,5 Hz y 40 Hz, eliminando los artefactos de baja frecuencia y la desviación de la línea base provocada por la respiración, así como el ruido de alta frecuencia causado principalmente por las interferencias de la red eléctrica. A continuación, eliminamos el primer y último segundo de cada señal para omitir la etapa de estabilización del filtrado.

Seguidamente, eliminamos los artefactos aberrantes de la señal, causados principalmente por movimientos bruscos y repentinos de los pacientes, que aparecían como cambios abruptos en el nivel de la señal. Para detectarlos, usamos una ventana deslizante de 0,5 segundos y calculamos sus valores máximo y mínimo. Si la diferencia estadística entre ventanas colindantes excedían determinados

umbrales máximos o mínimos, considerábamos anómalo dicho segmento y todos los valores de la señal dentro de esa ventana se establecían en cero.

Finalmente, eliminamos los segmentos de señal a partir de los primeros 15 segundos si existían. Por lo tanto, utilizamos segmentos cortos de ECG, de una duración máxima de 15 segundos, evitando así grandes diferencias en el tiempo de duración de las muestras del conjunto de datos, tal y como se hizo anteriormente en [Aublin 2021, Wickramasinghe 2021, Xiaoyu 2021].

5.3.4. Extracción de Características

Durante este estudio extrajimos automáticamente 81 características de cada una de las 12 derivaciones ECG disponibles en el conjunto de entrenamiento, además de la edad y el sexo de los metadatos de los registros, obteniendo un total de 974 valores de características por cada muestra. Estas características, la mayoría de ellas utilizadas anteriormente en [Zheng 2020a, Jiménez 2017a, Jiménez 2021], se derivaron de la actividad ventricular, basándose principalmente en la variabilidad del ritmo cardiaco, patrones de las ondas QRS y T, además de la parte baja del dominio espectral. A pesar de que parte de las variables extraídas dependían de la variabilidad del ritmo cardiaco, y esto debería proporcionar métricas similares para todas las derivaciones de ECG, se extrajeron de forma independiente para cada canal de ECG. Esto permitió tener en cuenta las diferencias en la detección del ritmo cardiaco para cada derivación, así como mantener la estructura del resto de variables dependientes de la derivación.

Para realizar esta tarea, en primer lugar, para cada derivación, extrajimos la secuencia *RR* utilizando nuestro detector de ondas QRS basado en el algoritmo de Pan y Tompkins y la primera derivada del ECG [Pan 1985]. A continuación, filtramos los valores atípicos de la secuencia *RR* utilizando la media \pm tres desviaciones estándar como umbrales y obtuvimos las derivadas primera y segunda de esa secuencia, denominadas aquí *RRd1* y *RRd2*. Además, creamos un detector de ondas T usando una ventana de 300 ms y un desplazamiento de 100 ms de las ondas R previamente identificadas, obteniendo y tratando el índice del valor absoluto máximo en esta ventana como la ubicación central de la onda T. Con esta información obtuvimos el intervalo QT y otras características relacionadas.

A continuación, obtuvimos ambos patrones de las ondas QRS y T en cada derivación, utilizando una ventana de \pm 100 ms sobre todas las detecciones de onda QRS y T anteriores.

Por otro lado, obtuvimos la estimación de la densidad espectral de potencia de Welch de cada derivación con el fin de calcular algunas características basadas en la frecuencia.

Utilizando las marcas de ondas QRS y T anteriores, las secuencias *RR*, los patrones de ondas y la información espectral, agrupamos en nueve categorías diferentes las 81 características de señal extraídas de cada derivación, las cuales se

detalladas en la Tabla 5.4. El buen funcionamiento de los detectores de ondas QRS y T está estrechamente relacionado al buen rendimiento de la clasificación final ya que 72 de estas 81 características dependen de tales marcas fiduciales.

Tabla 5.4. Detalle de las nueve categorías de características extraídas en cada derivación ECG.

Número y descripción de características en cada categoría. Además, la edad y el sexo se obtuvieron de los metadatos de los registros ECG.

Categoría	Nº	Descripción
1. Voltajes	4	Estadísticas básicas sobre los voltajes de las ondas R y T (media, desviación estándar).
2. Duración QT y QTc	5	Estadísticas básicas sobre el intervalo QT en milisegundos (media, desviación estándar). Duración media del QT corregido con las fórmulas de [Bazzet 1920], [Fridericia 2003] y [Sagie 1992].
3. Patrones QRS y T	25	Características basadas en los patrones QRS y T: Porcentaje de la amplitud de la onda T respecto a la onda R; signo de las ondas R y T (positivo o negativo); porcentaje de descartes de ondas y Error Cuadrático Medio (RMSE) durante la definición de los patrones R y T; valores máximos para la primera y segunda derivada de los patrones R y T. También dividimos los patrones en tres tramos t_i : t_1) antes del valor máximo absoluto; t_2) después del valor máximo absoluto y antes del valor mínimo absoluto; t_3) después del valor mínimo. Para cada tramo, obtuvimos los valores máximo y mínimo de la primera derivada. Además, obtuvimos un valor que indicaba si el máximo o mínimo era dominante en cada patrón.
4. Estadísticas <i>RR</i>	9	Estadísticas básicas sobre las secuencias <i>RR</i> , <i>RRd1</i> y <i>RRd2</i> seleccionando sólo aquellas estadísticas que presentaron diferencias significativas entre las muestras etiquetadas como positivas o negativas: media (<i>RR</i> , <i>RRd1</i>), desviación estándar (<i>RR</i> , <i>RRd1</i> , <i>RRd2</i>), curtosis

Categoría	Nº	Descripción (<i>RR</i> , <i>RRd1</i> , <i>RRd2</i>) y asimetría (<i>RR</i>).
5. Basadas en pNNx	4	Características complejas basadas en la secuencia <i>RRd1</i> : Media cuadrática de diferencias sucesivas (RMSSD). pNN25, pNN50, pNN75, donde pNNx indica el porcentaje de intervalos entre latidos normales que exceden los 25, 50 y 75 ms [Mietus 2002].
6. Gráfica de Poincaré	7	Características basadas en la gráfica de Poincaré utilizando la secuencia <i>RRd1</i> : Valores máximos, mínimos, media, desviación estándar, curtosis y asimetría de las distancias entre todos los puntos de la gráfica, así como la diferencia absoluta entre los valores de distancia máxima y mínima.
7. Gráfica de Lorenz	8	Características basadas en la gráfica de Lorenz utilizando la secuencia <i>RRd2</i> : Variabilidad angular: <i>Vector Angular Index</i> (VAI) y <i>Vector Length Index</i> (VLI) [Ruan 2011]. Dispersión de la distancia entre puntos al origen y diferencias entre 2 y 3 latidos consecutivos.
8. Características Espectrales	17	Frecuencia dominante (FD) utilizando el método de estimación de densidad espectral de Welch; concentración espectral en $FD \pm 0.5\text{Hz}$ en el periodograma normalizado en el rango [0, 1] y la suma del mismo periodograma normalizado en pasos de 2 Hz en el rango de frecuencias [0, 28] Hz.
9. Otras características	2	Complejidad de Lempel-Ziv de la serie temporal <i>RR</i> después de la binarización utilizando la mediana como umbral. Entropía de Shannon de la secuencia <i>RR</i> .
Número de características totales por derivación ECG: 81		

5.3.5. Preprocesado del Conjunto de datos

Una vez extraídas las características de los registros ECG descritas anteriormente, realizamos un preprocesado del conjunto de datos resultante. Para ello, en primer lugar, para cada característica en el pliegue de entrenamiento, los valores atípicos que excedían tres veces la desviación estándar por encima o por debajo de la mediana fueron reemplazados por estos mismos límites, saturando de este modo los valores anómalos en dichos extremos. A continuación, si alguna muestra contenía un valor *NaN* en alguna de sus características debido a un error de extracción o a la imposibilidad de obtener dicho valor, se reemplazó este valor *NaN* por la mediana en el conjunto de datos para dicha característica. De acuerdo con las dos reglas anteriores, el 1,24% de los valores resultaron ser atípicos, mientras que el 0,31% de los valores resultaron ser *NaN*, siendo sustituidos por su correspondiente valor límite o mediana.

Por último, para cada pliegue de entrenamiento durante la validación cruzada, realizamos una normalización *z-score* para reescalar el conjunto de datos, guardando la media (μ) y la desviación estándar (σ) de cada característica siguiendo la siguiente expresión

$$z = \frac{x - \mu}{\sigma} \quad (5.4)$$

donde x indica el valor de la característica original, mientras que z es el reescalado. Durante la etapa de validación de nuestra validación cruzada, usamos los valores guardados de μ y σ con el fin de volver a escalar de la misma manera los valores de las características que servirían de entrada para los modelos de clasificación. Además, en la etapa de validación, se reemplazaron los valores *NaN* por los valores de la mediana del pliegue de entrenamiento, puesto que los modelos de clasificación binaria no podían manejar valores *NaN*. Por otro lado, los valores atípicos que aparecían en el pliegue de validación no fueron reemplazados por ningún otro valor.

5.3.6. Selección de Características

Aplicamos un proceso de selección de características al conjunto de datos de entrenamiento utilizando métodos de filtrado estadísticos supervisados y no supervisados. Para ello, utilizamos como entrada las características correspondientes a las derivaciones ECG disponibles (ver Tabla 5.3). En primer lugar, la edad y el sexo siempre fueron incluidas entre las variables seleccionadas evitando así un posible conjunto vacío de características. A continuación, para cada característica realizamos un *test t de Student* para dos muestras con un valor α de 0,05, siendo la primera muestra aquellos valores de la característica que pertenecían a la clase

positiva, y la segunda muestra los valores que pertenecían a la clase negativa. Se eliminaron todas las características que no pasaron esta prueba de significación estadística. Hay que tener en cuenta que el *test t de Student* en grandes conjuntos de datos es muy sensible a las diferencias entre medidas, permitiendo identificar variables con potencial valor predictivo.

Cabe destacar que la selección de características anterior presentó la limitación de incluir muchas variables con información correlacionada. Por lo tanto, además de dicha selección, añadimos un paso adicional donde para cada posible par de características restantes, eliminamos una de ellas si su coeficiente de correlación de *Pearson* era mayor o igual a 0,95.

5.3.7. Estrategia de Clasificación Binaria

Para cada afección cardíaca a detectar (Tabla 5.2), se entrenó y validó un conjunto de clasificadores binarios para seleccionar finalmente aquel que presentase un mejor rendimiento de clasificación. El modelo seleccionado se incluyó en el sistema final de clasificación múltiple, descrito en la Sección 5.3.8, con el fin de etiquetar muestras como pertenecientes (muestra positiva) o no (muestra negativa) a cada una de las clases existentes en el conjunto de datos.

Diferenciamos dos tipos diferentes de clasificadores binarios entrenados y validados durante esta etapa: Clasificadores Supervisados (CSu) y Clasificadores Híbridos (CHi). Los primeros son modelos supervisados clásicos entrenados y validados con todo el conjunto de datos de entrenamiento. El otro modelo híbrido propuesto combina un algoritmo de *k-medias* no supervisado con el enfoque CSu anterior. Una vez validados todos los modelos de clasificación binaria para una determinada clase, se seleccionó el que presentó un valor mayor de puntuación G .

Asimismo, se realizó la selección de características descrita anteriormente para cada una de las 26 categorías distintas de ECG antes del entrenamiento de cada clasificador binario. De este modo, cada clasificador binario presentado a continuación tenía asociado un subconjunto de características seleccionadas específicamente para su propio proceso de entrenamiento y validación.

Las siguientes secciones detallan los procesos de entrenamiento para los modelos binarios CSu y CHi, además del proceso de selección del modelo binario final (CSu o CHi) con mejor rendimiento para cada afección cardíaca.

5.3.7.1. Entrenamiento de Clasificadores Supervisados (ECSu). El esquema del entrenamiento para un clasificador binario supervisado se muestra en la Figura 5.1. Allí, usamos como entrada el conjunto de datos X además del vector y_i de etiquetas de referencia (o *gold standard*) para una clase dada. A continuación, realizamos la selección de características obteniendo el subconjunto de datos X' , del que son eliminadas aquellas que no superan dicha selección.

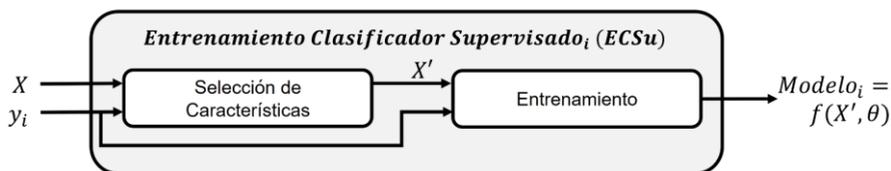


Figura 5.1. Diagrama resumen del proceso de Entrenamiento de un Clasificador Supervisado (ECSu).

Por último, realizamos un proceso de entrenamiento y validación para dos tipos distintos de modelos de clasificación: *Feed Forward Neural Networks* (FFNN) y *Naïve Bayes* (NB). A continuación, seleccionamos aquel modelo con un mayor valor de puntuación en la métrica G . Como resultado, se obtuvo una función de clasificación binaria llamada f , que utilizaba θ como parámetros del modelo (obtenidos durante el proceso de entrenamiento), y las características seleccionadas X' como las entradas necesarias.

En cuanto al modelo Naïve Bayes, usamos X' para entrenar un clasificador binario para cada clase usando la función *fitcnb* de Matlab R2021a. De forma predeterminada, esta función modela la distribución de la predicción de cada clase utilizando una distribución gaussiana que tiene una media y una desviación estándar; por otro lado, también modela la distribución de probabilidad *a priori* de la clase como una distribución de frecuencia relativa de las clases en el conjunto de datos.

Por otra parte, entrenamos dos redes FFNN diferentes, ambas con una arquitectura de una capa oculta y usando *tansig* como función de activación: la primera FFNN formada por 18 unidades en su capa oculta, mientras que la segunda por 32 unidades. Además, la capa de salida utilizó *mapminmax* como función de activación escalando la salida en el rango $[-1, 1]$. Aparte de esto, todas las FFNN fueron entrenadas con las funciones, objetos y parámetros predeterminados en *Matlab R2021a Deep Learning Toolbox*, usando *trainscg* (*Scaled Conjugate Gradient*) como función de entrenamiento. En este sentido, el flag *useGPU* se activó para usar las unidades de procesamiento gráfico disponibles con el fin de acelerar el entrenamiento. Finalmente, el flag *showResources* se desactivó para evitar la muestra de la interfaz gráfica de usuario que muestra el avance del entrenamiento pero que ralentiza considerablemente todo el proceso.

En cuanto la división del conjunto de datos de entrenamiento, se utilizó como entrada el 75 % de los datos en X' para entrenar las FFNN. Con el 25 % restante, seleccionamos para cada FFNN el umbral que maximizaba el valor de la métrica G para la unidad de salida dentro del rango $[-1, 1]$, usando para ello pasos de 0,005. Este umbral se utilizó como punto de corte para clasificar las muestras como positivas o negativas. De esta forma, lidiamos con el problema de las clases desbalanceadas en el conjunto de datos donde la proporción de muestras positivas para cada afección cardíaca siempre fue menor que las negativas.

Finalmente, de entre los tres clasificadores supervisados entrenados durante el proceso ECSu (un NB y dos FFNN distintas), elegimos el que presentó un mayor valor de la métrica de clasificación G , denominado aquí G_{ECSu} .

5.3.7.2. Entrenamiento de Clasificadores Híbridos (EChi). La Figura 5.2 muestra el enfoque de clasificación binaria híbrida propuesto en este estudio. Al igual que en la aproximación anterior, en primer lugar realizamos una selección de características mediante el conjunto de datos de entrada X y el vector y_i de etiquetas de referencia (o *gold standard*) para una clase dada, obteniendo como resultado el subconjunto de características seleccionadas X' . A continuación, el punto central de nuestra propuesta se basa en la técnica de aprendizaje automático no supervisado (k -medias) que divide en tres agrupaciones o *clústers* diferentes ($k = 3$) el conjunto de entrenamiento de entrada X' con el fin de obtener tres centroides distintos, uno para cada agrupación. Denominamos $C_{\{1,2,3\}}$ al vector que contiene estos centroides, y $X'_{\{1,2,3\}}$ a los tres subconjuntos con las muestras asociadas a cada centroide. Seguidamente, para cada subconjunto de muestras en $X'_{\{1,2,3\}}$, realizamos el proceso ECSu descrito anteriormente para obtener un clasificador de tipo CSu con el mejor rendimiento de clasificación posible en cada agrupación de datos. Finalmente, como salida de este proceso EChi, guardamos los índices de las características seleccionadas para cada agrupación en $X'_{\{1,2,3\}}$, los centroides del clúster $C_{\{1,2,3\}}$ y los tres clasificadores binarios CSu correspondientes a cada *clúster*, llamados aquí $f_{\{1,2,3\}}$.

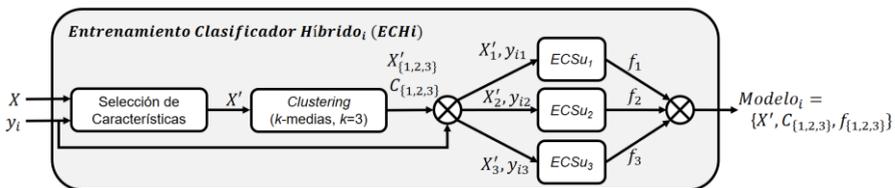


Figura 5.2. Diagrama resumen del proceso de Entrenamiento de un Clasificador Híbrido (EChi).

De este modo, utilizamos tres pasos para clasificar una nueva muestra no conocida denominada x . El primer paso es eliminar las características no seleccionadas de la tupla de valores dando como resultado la nueva tupla x' . El segundo es obtener la distancia euclídea de esta muestra x' a cada uno de los tres centroides en $C_{\{1,2,3\}}$, eligiendo el *clúster* que minimiza esta distancia. Finalmente, el tercer paso es utilizar el modelo CSu en $f_{\{1,2,3\}}$ asociado a dicho *clúster* para dar la respuesta binaria, es decir, usamos el modelo NB o FFNN entrenado previamente sólo con los datos pertenecientes a este *clúster* para obtener la etiqueta correspondiente a la muestra, denominando \hat{y}_i a la etiqueta predicha por nuestro sistema híbrido. Este proceso de clasificación se expresa de la siguiente manera,

$$\hat{y}_i = f_{\{c_{x'}\}}(x') \mid c_{x'} = \operatorname{argmin} \left(\text{distancia}(C_{\{1,2,3\}}, x') \right) \quad (5.5)$$

Cabe destacar que escogimos el valor arbitrario de $k = 3$ con el objetivo de obtener un compromiso entre el tiempo de entrenamiento y la mejor precisión de clasificación posible de los modelos resultantes. En este sentido, se debe tener en cuenta que el valor de k aumenta linealmente el número total de modelos CSu (n_{CSu}) a ser entrenados y evaluados en nuestro enfoque ECHi propuesto, siguiendo la siguiente expresión:

$$n_{CSu} = k \times n_{clases} \quad (5.6)$$

Además, la precisión de los clasificadores binarios supervisados asociados con cada *clúster* también dependía de la cantidad de muestras de entrada proporcionadas durante el proceso de entrenamiento. Dado que algunas clases tenían una ratio de muestras positivas muy pequeño, un valor de k superior a 3 resultaba en grupos con pocas muestras positivas, e incrementaban considerablemente el tiempo de entrenamiento, siendo el valor de 3 el óptimo de entre todos los evaluados. De esta manera, con 26 clases a identificar y un valor de $k = 3$, se entrenaron y validaron un total de 78 clasificadores de tipo CSu usando esta estrategia híbrida.

Una vez seleccionado el número de *clústers*, utilizamos los siguientes parámetros de entrenamiento para el algoritmo *k-medias*: un máximo de 200 iteraciones y 10 posibles repeticiones en cada una de ellas para tratar de lograr la máxima separación entre *clústers* y sus correspondientes centroides.

Por otro lado, para evaluar la precisión del modelo híbrido (G_{ECHi}), se tomó en cuenta la métrica G de cada clasificador supervisado asociado con un *clúster*, ponderando el número de muestras en dicho *clúster* con la siguiente expresión

$$G_{ECHi} = \sum_{i=1}^3 G(f_i) \left(\frac{n_i}{n} \right) \quad (5.7)$$

donde $G(f_i)$ y n_i denotan el valor de G y el número de muestras utilizadas para obtener el clasificador perteneciente al *clúster* i , mientras que n es el número total de muestras en el pliegue de entrenamiento. Por lo tanto, las agrupaciones grandes tuvieron mayor importancia en la puntuación G_{ECHi} del modelo, mientras que por el contrario, las agrupaciones de datos con menor número de muestras aportaron un menor peso a esta puntuación. Además, si algún grupo tenía menos de 40 muestras, o menos de 40 muestras positivas, no entrenamos ni validamos los clasificadores supervisados f_i correspondientes, dando un valor nulo a la puntuación $G(f_i)$ de este grupo, ya que no era viable alcanzar ni evaluar una clasificación binaria correcta con un número tan reducido de datos.

5.3.7.3. Selección de modelos. Una vez que entrenamos y validamos los enfoques de los modelos de clasificación CSu y CHi, seleccionamos sólo un clasificador binario con el fin de utilizarlo en el sistema de clasificación múltiple final, el cuál presentaremos en la siguiente sección 5.3.8. Para ello, elegimos el que alcanzó una mayor puntuación G , siguiendo la siguiente expresión

$$\text{Clasificador}_i = \text{Clasificador}(\text{argmax}(G_{ECSu}, G_{ECHi})) \quad (5.8)$$

donde G_{ECSu} y G_{ECHi} son los valores de puntuación G correspondientes a los modelos binarios seleccionados durante el proceso ECSu y nuestra propuesta híbrida ECHi.

5.3.8. Enfoque de Clasificación *One-vs-Rest*

Una vez que seleccionamos el clasificador binario que presentaba mejores resultados para cada una de las 26 categorías cardiacas a detectar, construimos un sistema final de clasificación múltiple que se ilustra en la Figura 5.3. Este enfoque también se conoce como modelo de clasificación Uno-Contra-Todos o *One-vs-Rest* por su traducción del inglés, donde las muestras del conjunto de datos se etiquetan como positivas o negativas utilizando cada clasificador binario para cada clase conocida. De esta manera, el modelo final de clasificación *One-vs-Rest* dará una respuesta binaria indicando si una muestra no vista anteriormente pertenece o no a cada una de las 26 clases contempladas en el proceso de entrenamiento. Además, como se detalló anteriormente, cada clasificador binario utiliza el conjunto seleccionado de características que mejor se adapta a su propio problema de clasificación. En consecuencia, cada modelo binario resuelve un problema de clasificación independiente en todo el sistema de clasificación, siendo posible asignar una nueva muestra a ninguna, una o más de una clase. Esto es especialmente útil en este estudio ya que algunas muestras podrían no pertenecer a ninguna categoría cardiaca previamente entrenada, o por el contrario pertenecer a más de una afección cardiaca.

Finalmente, hicimos una excepción durante el entrenamiento de la clase DEjD, donde siempre usamos un clasificador binario NB. La razón principal de esto fue que los enfoques de clasificación ECSu y ECHi dieron valores de puntuación G muy pequeños e inconsistentes entre los distintos pliegues de validación utilizando las redes FFNN o el clasificador binario híbrido para dicha afección. No obstante, NB presentó para este caso una puntuación G más consistente y elevada durante el proceso de validación, sin sobreajustar los resultados al conjunto de datos de entrada.

Enfoque de Clasificación *One-vs-Rest*

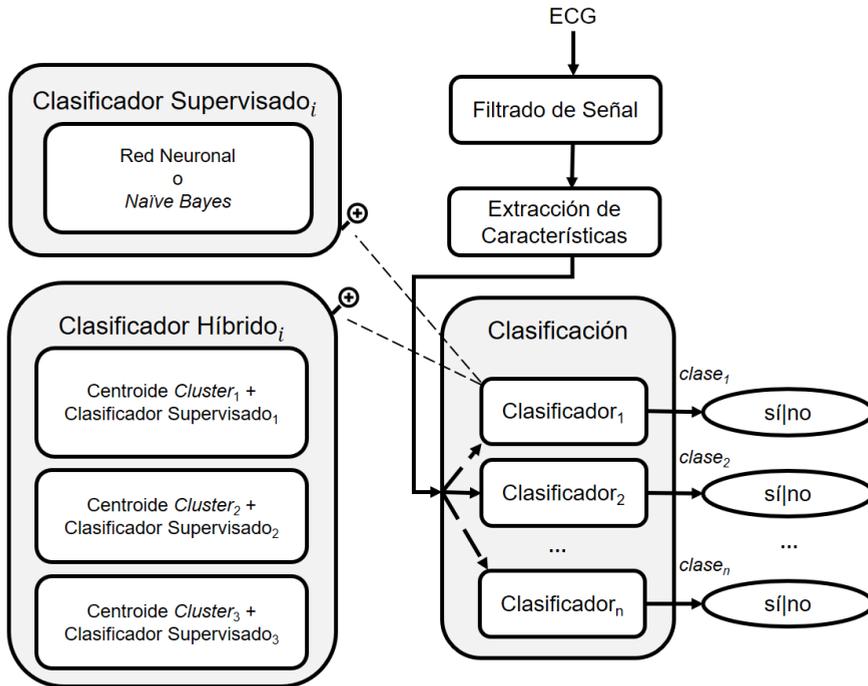


Figura 5.3. Diagrama resumen del enfoque de clasificación *One-vs-Rest* empleado

5.4. Resultados

5.4.1. Costes Computacionales

Cada uno de los tres pliegues utilizados durante la validación cruzada contenía aproximadamente 58.800 muestras para entrenamiento y 29.400 muestras para *test*. Con este número de muestras como contexto, cronometramos los tres procesos principales de nuestra experimentación: procesado y extracción de características de las señales, entrenamiento (selección de características incluida) y *testeo* de los modelos (tiempo necesario para clasificar muestras que los clasificadores no habían visto anteriormente).

El tiempo necesario para realizar el procesado de señal y la extracción de características para los registros de ECG de 12 derivaciones fue de $566,79 \pm 24,95$ segundos ($9,45 \pm 0,42$ minutos), lo que representa un rendimiento de $103,74 \pm 4,16$ muestras por segundo utilizando el hardware descrito anteriormente.

A continuación, medimos el tiempo necesario para realizar el entrenamiento, validación y selección de los multclasificadores *One-vs-Rest* correspondientes a cada combinación de conjuntos de derivaciones ECG. Finalmente, realizamos la misma operación para cronometrar el proceso de *testeo* donde los modelos clasifican las muestras no vistas anteriormente. La Figura 5.4 muestra el tiempo medio de cómputo necesario para ambos procesos, siendo el entrenamiento el proceso que más tiempo consume con una media de 3,4 horas para 12 derivaciones y 46 minutos para una. Por otra parte, que el *testeo* presentó un comportamiento lineal en el tiempo de cómputo requerido con una pequeña pendiente con respecto al número de derivaciones, utilizando una media de 66 minutos para 12 derivaciones y 27 minutos para una derivación.

Finalmente, la Figura 5.5 muestra el rendimiento durante el entrenamiento y *testeo* de los modelos en términos del número de muestras procesadas por segundo. El mejor rendimiento corresponde al clasificador de una sola derivación ECG, con $21,20 \pm 0,35$ muestras/segundo durante el entrenamiento y $18,02 \pm 0,18$ muestras/segundo durante el *testeo*. Como era de esperar, se obtienen resultados de rendimiento más bajos con el clasificador de ECG de 12 derivaciones, con $4,86 \pm 0,02$ muestras/segundo para el entrenamiento y $7,46 \pm 0,07$ muestras/segundo durante el *testeo*. Debe tenerse en cuenta que cada vez que se clasificaba una nueva muestra en el proceso de *test*, la competición requería de la lectura del archivo de dicha muestra del disco, lo que agregaba una sobrecarga computacional que no existía durante el proceso de entrenamiento.

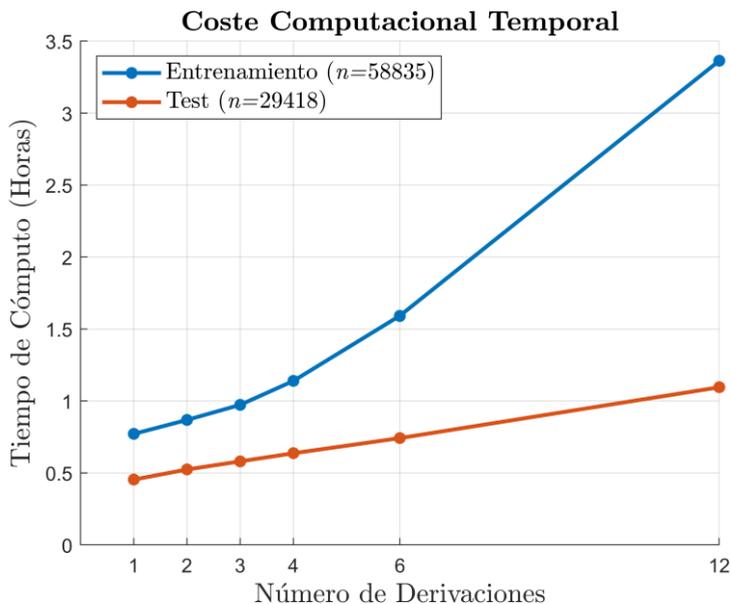


Figura 5.4. Tendencia del tiempo medio de entrenamiento y *test* de cada multclasificador de ECG durante la validación cruzada triple en el conjunto de datos público para entrenamiento.

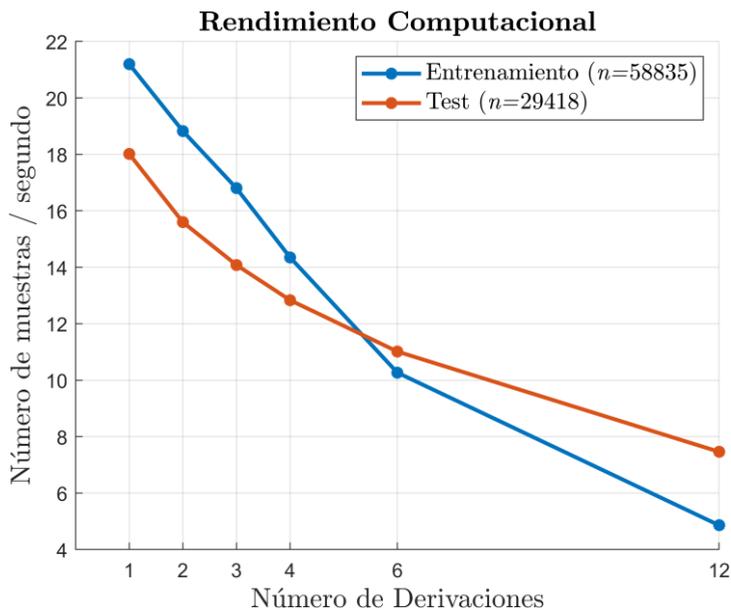


Figura 5.5. Rendimiento medio de cada multclasificador de ECG durante la validación cruzada triple en el conjunto de datos público para entrenamiento, en términos de número de muestras procesadas por segundo.

5.4.2. Análisis de la Selección de Características

En primer lugar, la Tabla 5.5 muestra el porcentaje de características seleccionadas de todo el conjunto de datos con el objetivo de entrenar los clasificadores binarios que finalmente componían los modelos multclasificadores *One-Vs-Rest* para cada combinación de conjunto de derivaciones ECG. El método empleado seleccionó una media del $60\pm 10\%$ de las características disponibles (eliminando el resto para el clasificador binario correspondiente) sin presentar diferencias significativas entre las combinaciones de derivaciones utilizadas.

En segundo, para conocer el rendimiento y comportamiento del método de selección entre las diferentes derivaciones de ECG, obtuvimos el porcentaje de las características seleccionadas para cada clasificador binario y las agrupamos por la derivación ECG de procedencia; estos resultados se muestran en la Tabla 5.6. Durante el entrenamiento de los modelos con 12 derivaciones, aunque las características extraídas de la derivación V1 presentaron un porcentaje de selección ligeramente mayor con un valor de $8.81\pm 0.76\%$, no existían diferencias significativas en este sentido. En el caso de los modelos entrenados con 4 y 3 derivaciones, las características extraídas de la derivación V2 presentaron un menor porcentaje de selección con valores de $24,15\pm 2,28\%$ y $32,01\pm 2,72\%$ cada una. Las demás combinaciones de derivaciones tampoco presentaron diferencias significativas en los porcentajes entre características correspondientes a distintas derivaciones, siendo esta relación proporcional al número de derivaciones empleadas.

Resumiendo los resultados presentados en las Tabla 5.5 y 5.6, se observó un filtrado medio del $40\pm 10\%$ de las características extraídas utilizando nuestro método de selección y que la proporción de estas características se equilibró entre las derivaciones ECG disponibles en cada modelo de clasificación binaria. De este modo, fuimos capaces de aliviar la carga computacional que supondría el entrenamiento y validación de los modelos de clasificación con el 100% de las características disponibles.

En tercer y último lugar, se obtuvo el porcentaje de características seleccionadas agrupadas por categoría (ver Tabla 5.7). La categoría más relevante corresponde a las características relacionadas con la duración del intervalo QT (Gr1) con una media de selección del 86%, y las estadísticas de los voltajes de las ondas R y T (Gr2) con una media de selección del 75%. La Figura 5.6 ilustra esta circunstancia en diagramas de cajas, donde se aprecia que los dos primeros grupos tienen en general mayor relevancia en la selección. No obstante, la media de selección del resto de categorías de características se situó entre el 67% y el 43%, no encontrándose ninguna característica totalmente irrelevante en toda la experimentación. Cabe destacar que las características pertenecientes al Gr1 se seleccionaron el 100% de las veces para los clasificadores binarios correspondientes a las patologías DEjI, MPrR y FLA.

Tabla 5.5. Análisis de la selección de características agrupado por conjuntos de derivaciones.

Porcentaje (media±desviación estándar) de las características seleccionadas para cada clasificador binario durante la validación cruzada triple en el conjunto de datos de entrenamiento público.

Nº de Derivaciones	Características Seleccionadas [%]
12	59,6±10,0
6	61,0±10,3
4	60,2±10,2
3	60,4±9,9
2	61,4±10,9
1	61,6±10,5

Tabla 5.6. Análisis de la selección de características agrupado por derivaciones.

Porcentaje ($\mu \pm \sigma$) de las características seleccionadas para los clasificadores binarios utilizados durante la validación cruzada triple, agrupados por derivaciones ECG. La edad y el sexo siempre fueron seleccionados.

[%]	12 deriv.	6 deriv.	4 deriv.	3 deriv.	2 deriv.
I	8,48±0,67	16,55±1,25	25,05±1,79	33,22±2,18	48,86±2,61
II	8,50±0,57	16,60±0,90	25,15±1,68	33,37±2,33	49,07±2,75
III	8,32±0,72	16,25±1,34	24,59±1,86	-	-
aVR	8,57±0,61	16,74±1,03	-	-	-
aVL	8,69±0,72	16,96±1,20	-	-	-
aVF	8,30±0,81	16,20±1,44	-	-	-
V1	8,81±0,76	-	-	-	-
V2	8,06±0,73	-	24,15±2,28	32,01±2,72	-
V2	7,98±0,62	-	-	-	-
V4	8,06±0,50	-	-	-	-
V5	7,86±0,44	-	-	-	-
V6	8,01±0,42	-	-	-	-

Tabla 5.7. Análisis de la selección de características agrupado por categorías. Porcentaje ($\mu \pm \sigma$) de las características seleccionadas para cada clasificador binario durante la validación cruzada triple en el conjunto de datos de entrenamiento público.

Grupos de Características	Características Seleccionadas [%]
1 - Voltaje (R, T)	74,9±11,2
2 - Duracion QT y QTc	86,0±12,7
3 - Patrones QRS y T	57,8±7,7
4 - Estadísticas RR	62,1±18,7
5 - Basadas en pNNx	60,8±13,5
6 - Gráfica de <i>Poincaré</i>	43,4±18,6
7 - Gráfica de Lorenz	54,1±17,5
8 - Características Espectrales	59,9±7,6
9 - Otras	66,2±30,2

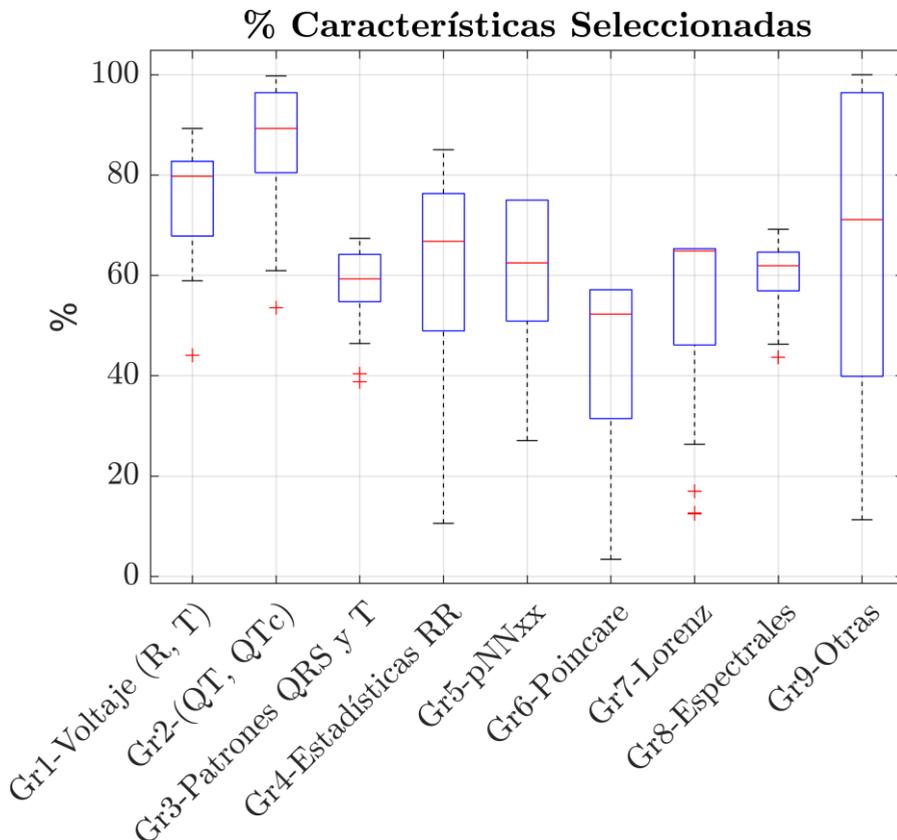


Figura 5.6. Diagramas de cajas del porcentaje de características seleccionadas en cada categoría/grupo (Gr).

5.4.3. Análisis de la Selección de Modelos

Los porcentajes de los distintos tipos de modelos seleccionados para los clasificadores binarios en función de las distintas combinaciones de derivaciones ECG utilizadas para el entrenamiento se muestran en la Tabla 5.8. En todas estas combinaciones de derivaciones, en más del 50% de los casos se seleccionó un modelo FFNN, mientras que en un rango entre el 28% y el 38% de los casos, nuestro clasificador híbrido propuesto presentó el mejor rendimiento, siendo seleccionado como el clasificador binario final para determinadas afecciones cardiacas. Por último, NB se seleccionó en menos del 11 % de los casos, principalmente cuando los otros modelos no funcionaron adecuadamente.

Asimismo, en la Tabla 5.9 se detalla el porcentaje de los distintos tipos de modelos seleccionados para cada clasificador binario. Para diez afecciones cardiacas el único modelo de clasificación seleccionado fue una red FFNN, mientras que, para las otras clases, se seleccionó una combinación de modelos dependiendo del pliegue de datos en la validación cruzada y/o el número de derivaciones ECG utilizadas para el entrenamiento. NB se usó en el 100% de las veces únicamente en el caso de la afección cardiaca DEjD, ya que usamos una regla definida previamente a tal efecto.

Cabe destacar que no se encontró correlación alguna entre el número de muestras positivas de cada afección cardiaca, y el tipo de modelo binario seleccionado, lo que indica que no existió sesgo en este sentido.

Tabla 5.8. Análisis de la selección de modelos agrupado por conjuntos de derivaciones.

Porcentaje de los distintos tipos de modelos seleccionados para los clasificadores binarios en función de la combinación de derivaciones ECG utilizadas para el entrenamiento durante la validación cruzada triple en el conjunto de datos de entrenamiento público.

Nº Derivaciones	Híbrido [%]	FFNN [%]	NB [%]
12	38,5	57,6	3,9
6	34,6	55,1	10,3
4	33,3	59,0	7,7
3	30,8	62,8	6,4
2	28,2	64,1	7,7
1	32,0	59,0	9,0

Tabla 5.9. Análisis de la selección de modelos agrupado por afecciones cardiacas.

Porcentaje de los distintos tipos de modelos seleccionados para cada clasificador binario durante la validación cruzada triple en el conjunto de datos de entrenamiento público con todas las combinaciones de derivaciones de ECG.

Clase	Híbrido [%]	FFNN [%]	NB [%]
AS	-	100	-
BAVPG	33,3	66,7	-
BFA _n I	11,1	88,9	-
BR	55,6	-	44,4
Bradi	-	88,9	11,1
BradiS	-	100	-
BRDC BRD	-	100	-
BRDI _n	94,4	5,6	-
BRIC BRI	83,3	-	16,7
CAP LSP	-	100	-
CVP LVP	-	100	-
DE _j D	-	-	100
DE _j I	50,0	50,0	-
FA	-	100	-
FLA	-	100	-
MPrP	94,4	-	5,6
PRPr	11,1	72,2	16,7
QAn	44,4	55,6	-
QTPr	100	-	-
RMp	94,4	5,6	-
RSN	66,7	33,3	-
TacS	-	100	-
Tan	-	100	-
TCII	33,3	66,7	-
TInv	-	100	-
VQRSB	83,3	16,7	-

5.4.4. Análisis de las Puntuaciones de los Modelos

Los mejores resultados usando la métrica de puntuación *Challenge Metric* (*CM*) en todo el conjunto oficial de *test* oculto mostraron un valor de 0,388 usando 12 derivaciones. Sin embargo, la puntuación más baja obtenida en el mismo conjunto completo de *test* tuvo como resultado un valor de *CM* de 0,373 usando sólo dos derivaciones ECG, con una ligera diferencia de 0,015 con respecto a la primera. Por otro lado, durante la validación cruzada triple en el conjunto de datos de

entrenamiento público, obtuvimos un valor de CM de $0,435 \pm 0,009$ usando 12 derivaciones y $0,388 \pm 0,005$ usando una única derivación ECG. No obstante, las diferencias entre las configuraciones de 6, 4, 3 y 2 derivaciones también fueron menores comparando sus valores de CM con la combinación de 12 derivaciones, siendo 0,033 la máxima diferencia encontrada. La Tabla 5.10 muestra el conjunto completo de resultados utilizando la métrica de puntuación CM durante nuestra validación cruzada además de los obtenidos en los diferentes conjuntos de datos para *test* ocultos. Los valores medios de AUROC en el conjunto de datos de *test* fueron de 0,86 en CPSC, 0,81 en G12EC, 0,84 en la base de datos *undisclosed* y 0,82 en el conjunto de pruebas UMich, sin diferencias significativas entre las distintas combinaciones de derivaciones.

La Tabla 5.11 muestra la media y la desviación estándar de diferentes métricas del rendimiento en la clasificación de las 26 clases puntuadas en la competición durante la validación cruzada en el conjunto de entrenamiento público, donde se logró un valor de G ligeramente superior a 0,80 utilizando 12 derivaciones, seguido de un valor G de 0,79 con 4 y 3 derivaciones y de 0,78 con 6 y 2 derivaciones. El modelo de clasificación para la derivación única obtuvo una puntuación G de 0,74 mostrando una relación equilibrada entre sensibilidad y especificidad de 0,75 y 0,76 respectivamente. La Figura 5.7 muestra estos valores de G agrupados por conjuntos de derivaciones en diagramas de caja, donde se aprecia que la tendencia no presenta diferencias significativas entre grupos.

En la Tabla 5.12, mostramos los resultados obtenidos para los clasificadores binarios individuales durante la validación cruzada en el conjunto de entrenamiento público, donde las siguientes diez afecciones cardíacas presentaron puntuaciones en la métrica G iguales o superiores a 0,85 en una o varias combinaciones de derivaciones ECG: FA, FLA, BRIC|BRI, BRDC|BRD, DEjI, BFAnI, RMp, AS, BradiS y TacS.

Finalmente, la Figura 5.8 muestra los diagramas de caja de los valores de puntuación G para cada una de las afecciones cardíacas durante la validación cruzada en el conjunto de entrenamiento público, utilizando los valores obtenidos con todas las combinaciones de derivaciones. Allí observamos que FA, FLA, BRIC|BRI, BRDC|BRD, RMp, AS, BradiS y TacS presentan valores en la puntuación G iguales o superiores a 0,83 en todas las combinaciones de derivaciones, sin diferencias significativas entre ellas. Esta condición también se cumple en DEjI y BFAnI excepto en el caso de utilizar una única derivación, donde el rendimiento cae en ambas afecciones a un valor de G de 0,66, lo que sugiere que aquí es conveniente usar siempre más de una derivación.

Tabla 5.10. Resultados de clasificación en la métrica *CM*.

Valores obtenidos de la métrica de puntuación *Challenge Metric (CM)* de la competición para nuestros modelos finales de clasificación de múltiples derivaciones usando validación cruzada triple en el conjunto de entrenamiento público, puntuación única en el conjunto de validación oficial y oculto, y puntuación única en cada conjunto de test oficial y oculto, así como en todo el conjunto de test oculto oficial. Los organizadores de la competición no evaluaron la configuración de una única derivación ECG.

Nº Derivaciones	Conjunto de Entrenamiento (<i>CM</i>)	Conjunto de validación (<i>CM</i>)	Conjunto de <i>test</i> CPSC (<i>CM</i>)	Conjunto de <i>test</i> G12EC (<i>CM</i>)	Conjunto de <i>test undisclosed</i> (<i>CM</i>)	Conjunto de <i>test</i> UMich (<i>CM</i>)	Conjunto de <i>test</i> completo (<i>CM</i>)
12	0,435±0,009	0,440	0,301	0,465	0,284	0,418	0,388
6	0,402±0,003	0,431	0,281	0,457	0,262	0,410	0,376
4	0,421±0,001	0,435	0,279	0,457	0,286	0,418	0,387
3	0,420±0,004	0,432	0,278	0,457	0,282	0,415	0,384
2	0,414±0,005	0,428	0,268	0,459	0,252	0,407	0,373
1	0,388±0,005	-	-	-	-	-	-

Tabla 5.11. Resultados de clasificación (Sensibilidad, Especificidad, *G*, AUROC) agrupados por conjuntos de derivaciones.

Media±desviación estándar de otras métricas de rendimiento en la clasificación de las 26 clases puntuadas para nuestros modelos finales de clasificación de derivaciones múltiples usando la validación cruzada triple en el conjunto de entrenamiento público

Nº Derivaciones	Sensibilidad	Especificidad	<i>G</i>	AUROC
12	0,817±0,003	0,795±0,007	0,804±0,004	0,817±0,008
6	0,784±0,005	0,779±0,005	0,777±0,001	0,811±0,017
4	0,800±0,005	0,787±0,002	0,788±0,002	0,828±0,014
3	0,801±0,002	0,784±0,004	0,786±0,002	0,837±0,009
2	0,787±0,004	0,779±0,004	0,775±0,002	0,823±0,011
1	0,751±0,006	0,757±0,006	0,744±0,005	0,784±0,007

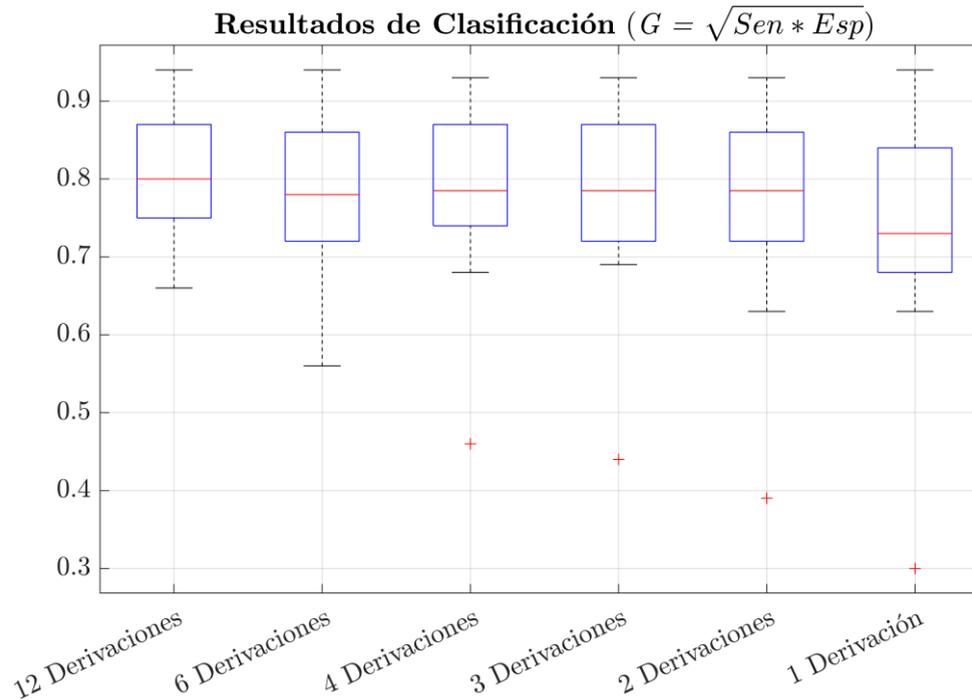


Figura 5.7. Diagramas de caja de los valores de la métrica G de los clasificadores binarios, durante la validación cruzada triple en el conjunto de datos de entrenamiento público, agrupados por cada conjunto de derivaciones.

Tabla 5.12. Resultados de clasificación (G) agrupados por conjuntos de derivaciones y afección cardiaca.

Valores de la métrica G de los clasificadores binarios para cada afección cardiaca, durante la validación cruzada triple en el conjunto de datos de entrenamiento público.

Clase	G (12 deriv.)	G (6 deriv.)	G (4 deriv.)	G (3 deriv.)	G (2 deriv.)	G (1 deriv.)
TacS	0,94±0,00	0,94±0,00	0,93±0,00	0,93±0,00	0,93±0,00	0,94±0,00
BradiS	0,93±0,00	0,93±0,00	0,93±0,00	0,93±0,00	0,93±0,00	0,93±0,00
BRIC BRI	0,91±0,01	0,90±0,00	0,90±0,01	0,89±0,01	0,88±0,01	0,87±0,02
BRDC BRD	0,91±0,01	0,85±0,00	0,88±0,01	0,88±0,01	0,86±0,01	0,85±0,00
BFA _{nI}	0,90±0,01	0,91±0,00	0,90±0,00	0,90±0,00	0,91±0,01	0,66±0,02
RMp	0,89±0,01	0,86±0,00	0,88±0,02	0,88±0,00	0,87±0,01	0,84±0,00
FLA	0,87±0,01	0,86±0,01	0,87±0,00	0,86±0,00	0,85±0,01	0,84±0,00
DEjI	0,87±0,00	0,86±0,00	0,83±0,02	0,83±0,00	0,83±0,00	0,66±0,00
FA	0,86±0,00	0,87±0,01	0,86±0,00	0,87±0,00	0,87±0,01	0,87±0,00
AS	0,84±0,01	0,83±0,01	0,84±0,00	0,85±0,00	0,84±0,01	0,86±0,01
CAP LSP	0,80±0,01	0,80±0,00	0,79±0,01	0,79±0,01	0,80±0,00	0,78±0,00
RSN	0,80±0,01	0,79±0,00	0,79±0,01	0,79±0,01	0,80±0,00	0,79±0,01
CVPLVP	0,80±0,03	0,78±0,00	0,79±0,01	0,79±0,02	0,78±0,02	0,76±0,01
BRD _{In}	0,80±0,01	0,69±0,01	0,75±0,01	0,75±0,01	0,68±0,01	0,66±0,01
VQRSB	0,79±0,01	0,76±0,00	0,76±0,01	0,78±0,01	0,77±0,01	0,70±0,01
MPrP	0,77±0,03	0,61±0,04	0,74±0,00	0,76±0,01	0,67±0,00	0,64±0,00
QTP _r	0,76±0,01	0,74±0,02	0,75±0,03	0,76±0,00	0,74±0,00	0,73±0,01
BR	0,76±0,01	0,73±0,04	0,75±0,03	0,72±0,00	0,72±0,02	0,68±0,01

Clase	G (12 deriv.)	G (6 deriv.)	G (4 deriv.)	G (3 deriv.)	G (2 deriv.)	G (1 deriv.)
TInv	0,76±0,00	0,72±0,01	0,74±0,00	0,72±0,01	0,72±0,01	0,70±0,01
BAVPG	0,75±0,03	0,78±0,01	0,76±0,00	0,77±0,01	0,78±0,00	0,75±0,00
Bradi	0,75±0,04	0,73±0,03	0,78±0,01	0,78±0,02	0,79±0,05	0,73±0,03
TAn	0,75±0,00	0,72±0,00	0,72±0,00	0,72±0,01	0,72±0,00	0,69±0,00
QAn	0,69±0,01	0,68±0,01	0,68±0,01	0,69±0,01	0,63±0,02	0,63±0,01
TCII	0,69±0,01	0,66±0,01	0,68±0,00	0,69±0,00	0,69±0,00	0,68±0,01
PRPr	0,66±0,03	0,66±0,02	0,71±0,02	0,70±0,00	0,69±0,02	0,69±0,03
DEjD	0,66±0,01	0,56±0,00	0,46±0,01	0,44±0,02	0,39±0,02	0,30±0,02

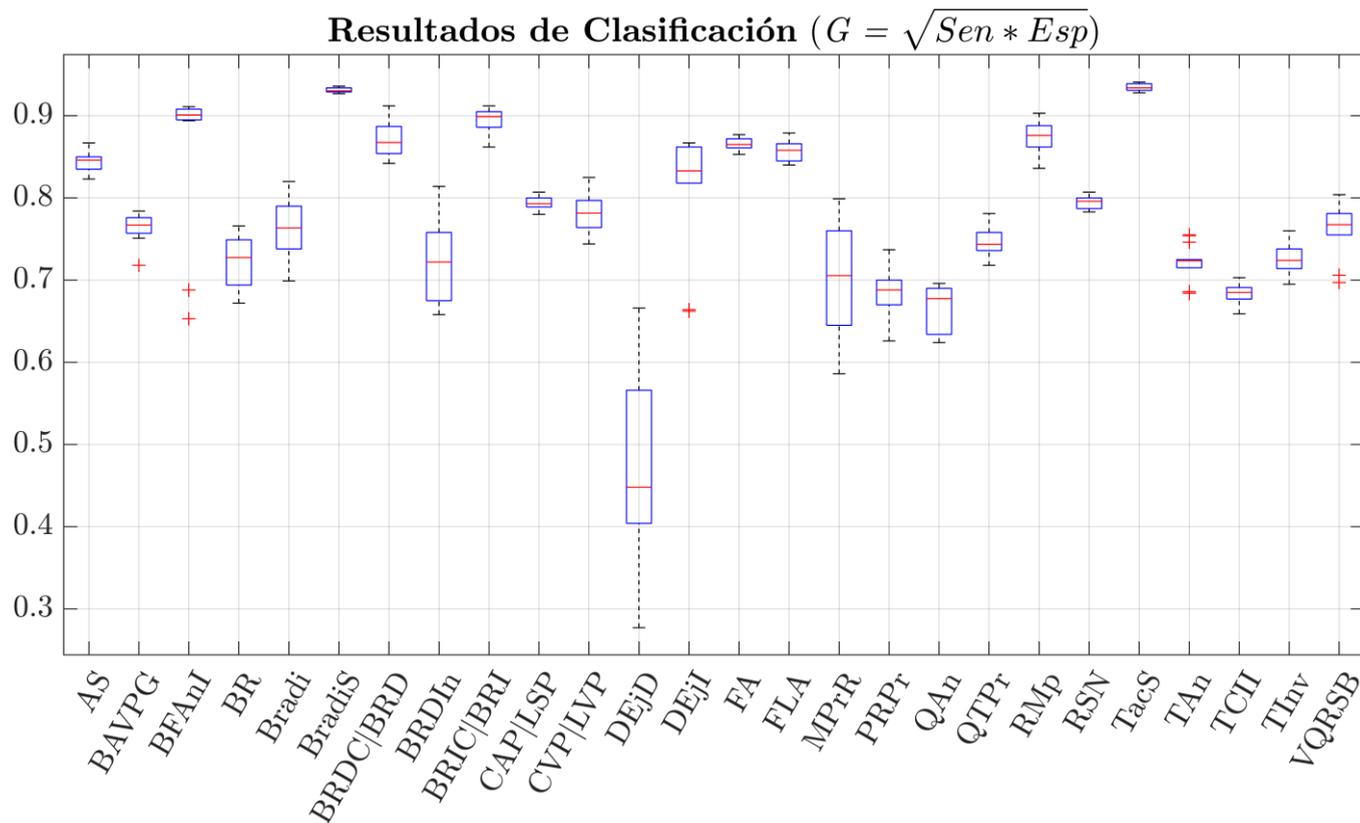


Figura 5.8. Diagramas de caja de los valores de la métrica G de cada clasificador binario, utilizando los valores obtenidos con todas las combinaciones de derivaciones ECG durante la validación cruzada triple en el conjunto de datos de entrenamiento público.

5.5. Discusión

Como era de esperar, el mejor rendimiento se alcanzó utilizando 12 derivaciones, lo que sugiere que este estándar no debe reemplazarse por sistemas de derivaciones reducidos en la práctica clínica. Sin embargo, la disminución del rendimiento con menos derivaciones fue pequeña, según los resultados obtenidos de las puntuaciones *CM* y *G*. Estos resultados muestran un gran potencial para conjuntos reducidos de derivaciones ECG fuera del entorno clínico, como por ejemplo, aquellos destinados a cribados y seguimientos masivos para la detección temprana de anomalías.

Por otro lado, cabe destacar que las clasificaciones que utilizaron 3 y 4 derivaciones superaron en rendimiento a aquellas que usaron 6. Aunque a primera vista parezca sorprendente, estos resultados pueden explicarse por el hecho de que, en el subconjunto de 6 derivaciones, únicamente dos de ellas son independientes; de hecho, el sistema de 6 derivaciones se registra utilizando sólo tres electrodos.

Asimismo, los subconjuntos de 3 y 4 derivaciones contienen la derivación precordial V2, la cual es independiente de las derivaciones de las extremidades y, por lo tanto, agrega información clínicamente valiosa. De acuerdo con esto, el rendimiento del subconjunto de 6 derivaciones debe contrastarse con el de 2 derivaciones para una comparación más justa. De hecho, los modelos con 2 derivaciones presentaron los resultados de puntuación más cercanos a los obtenidos con 6, donde obtuvimos valores de *CM* de 0,373 y 0,376 en el conjunto de *test* oculto, y valores medios de *G* de 0,775 y 0,777 respectivamente durante la validación cruzada triple. Estos resultados sugieren que es más importante incluir derivaciones con información complementaria entre sí que aumentar el número de ellas.

5.5.1. Resultados con una única derivación e idoneidad para dispositivos inteligentes

Centrándonos en los resultados obtenidos utilizando únicamente la derivación I durante la validación cruzada triple, se consiguieron valores medios de *CM* y *G* de 0,39 y 0,74, mientras que usando el conjunto completo de 12 derivaciones, las mismas puntuaciones fueron 0,43 y 0,80, respectivamente. Estos resultados son de gran interés, ya que utilizar una única derivación posibilita el registro de ECG y el diagnóstico automatizado en diversas situaciones diarias donde el estándar de 12 derivaciones no es viable o posible, a pesar del hecho de que el rendimiento es menor con una derivación. Esto es así, por ejemplo, usando un dispositivo portátil o un reloj inteligente, manipulando un teléfono móvil o mientras se conduce con un volante inteligente, por mencionar algunos casos.

Además, cabe destacar que, para seis afecciones cardiacas, el rendimiento de la clasificación utilizando únicamente la derivación I igualó o incluso mejoró la

clasificación con 12 derivaciones. En concreto, los casos donde los valores de G para clasificadores binarios que utilizaban una única derivación igualaron o mejoraron aquellos que utilizaban 12 durante la validación cruzada triple fueron los siguientes: 0,86 frente a 0,84 para AS; 0,87 frente a 0,86 para FA; 0,69 frente a 0,66 para PRPr; 0,75 con 1 y 12 derivaciones para BAVPG; 0,93 con 1 y 12 derivaciones para BradiS; y 0,94 con 1 y 12 derivaciones para TacS. La mayoría de estos resultados de clasificación binaria para una única afección cardiaca incluso mejoran utilizando el modelo de clasificación de 2 derivaciones ECG.

Sin embargo, este estudio también mostró que algunas afecciones cardiacas necesitan información de derivaciones precordiales para mejorar su detección, como BR, DEjI, BFAñI, VQRSB, MPrR y DEjD, donde la configuración de 12 derivaciones superó claramente a otras configuraciones. Esto sugiere que nuestro enfoque debe usarse con precaución cuando se pretende detectar afecciones cardiacas con bajo rendimiento en los resultados.

Desde un punto de vista computacional, el tiempo medio de clasificación con una única derivación ECG redujo a la mitad el tiempo medio de clasificación con 12 derivaciones durante la etapa de *test*. Por lo tanto, los dispositivos portátiles e inteligentes saldrían beneficiados de unos menores costes computacionales y de consumo de batería.

5.5.2. Mejoras respecto el trabajo reportado en *Computing in Cardiology 2021*

Una versión inicial de este trabajo fue previamente publicada en las actas del congreso *Computing in Cardiology 2021* en el contexto de la competición *Physionet Challenge* [Jiménez 2021]. Sin embargo, el estudio actual, reportado también en [Jiménez 2022] en el contexto del *Special Issue* titulado *Classification of Multilead ECGs* de la revista *Physiological Measurement*, presenta cambios importantes que nos condujeron a mejoras en el rendimiento de todas las métricas de puntuación. En cuanto a la puntuación CM , en [Jiménez 2021] obtuvimos valores de 0,34, 0,34, 0,27, 0,30 y 0,34 utilizando las combinaciones de 12, 6, 4, 3 y 2 derivaciones en el conjunto oficial de *test* oculto; mientras que en [Jiménez 2022] mejoramos las mismas puntuaciones con valores de 0,39, 0,38, 0,39, 0,38 y 0,37, respectivamente. Lo mismo sucedió con la puntuación de la métrica G durante la validación cruzada triple en el conjunto de entrenamiento público, obteniendo en [Jiménez 2021] valores de 0,76, 0,74, 0,69, 0,69, 0,74 y 0,71 usando combinaciones de 12, 6, 4, 3, 2 y 1 derivaciones ECG, mientras que [Jiménez 2022] mejoró las puntuaciones anteriores con valores de 0,80, 0,78, 0,79, 0,79, 0,78 y 0,74, respectivamente. Los cambios que realizamos para conseguir estas mejoras son los siguientes:

Después de la etapa de filtrado de la señal, usamos sólo 15 segundos de registro ECG, como se hizo en otros trabajos [Xiaoyu 2021, Wickramasinghe 2021, Aublin 2021], mejorando el tiempo de procesado y evitando así algunos efectos colaterales

no deseados relacionados con registros de larga duración. El principal problema en este sentido es que un registro largo puntuaba igual que uno corto, y en la base de datos la mayoría de registros ECG utilizados son de corta duración. Otro problema que encontramos con los registros largos fue que la probabilidad de encontrar un número relativamente elevado de afecciones cardíacas aumentaba, pero las etiquetas de los registros podrían no corresponderse en diferentes tramos de las señales. Por ejemplo, una fibrilación auricular paroxística podría aparecer y desaparecer en el mismo registro ECG de larga duración. De esta manera, dividir la señal en diferentes partes con el mismo etiquetado o usar una gran ventana deslizante durante la extracción de características podría convertirse en una operación sin sentido al entrenar los clasificadores.

Por lo tanto, es muy probable que la clasificación automática de registros ECG cortos de no más de 30 segundos deba abordarse de manera diferente a la clasificación de registros ECG más largos, así como el etiquetado de cada segmento de señal para registros largos. Recordemos que uno de los objetivos de este trabajo es conseguir modelos de clasificación adecuados para dispositivos inteligentes y capaces de manejar señales ECG cortas. Por ello, la metodología empleada evaluó las puntuaciones de clasificación en registros que no duraban más de 15 segundos, ya que en la etapa de procesado de señal fueron recortados si excedían esa duración. Además, una diferencia más con respecto a [Jiménez 2021] es que en este trabajo ya no usamos características específicas para registros de ECG largos, como lo son las estadísticas extraídas utilizando una ventana deslizante sobre señales largas. De este modo, en total eliminamos 46 características de este tipo utilizadas previamente. Por otro lado, añadimos características más específicas relativas al patrón QRS y T que mejoraron nuestra puntuación de clasificación.

A continuación, añadimos el preprocesamiento de datos para usar sólo 26 clases mezclando las afecciones cardíacas que ponderaban igual, y también evitamos usar muestras que no pertenecían a ninguna de las 26 categorías cardíacas para el entrenamiento de modelos. También cambiamos el valor del umbral en la segunda etapa de la selección de características, pasando de 0,90 a 0,95 en el coeficiente de correlación entre dos características diferentes, mejorando también así los resultados ligeramente.

Finalmente, agregamos un enfoque mixto utilizando técnicas de aprendizaje automático supervisadas y no supervisadas, donde cada clasificador binario podría estar compuesto de un clasificador FFNN, NB o híbrido, mientras que en [Jiménez 2021] únicamente usamos FFNN como clasificadores binarios. El sistema de clasificación híbrido propuesto se basa en un algoritmo *k-medias* no supervisado, donde se buscan 3 *clústers* y sus centroides, y se asocia una FFNN o NB a cada uno de ellos. Además, se creó un sistema de selección de modelos de estos clasificadores binarios para el proceso de entrenamiento.

En último lugar, solucionamos algunos errores e imprecisiones menores en el código, publicando la última versión de éste en

https://github.com/sjimenezupv/itaca_upv.cinc2021.special_issue. Como resultado, mejoramos todas las puntuaciones de clasificación del trabajo inicial.

5.5.3. Comparación con otros trabajos

Durante la competición del *Computing in Cardiology 2021*, se propusieron otros trabajos para abordar este problema de clasificación para múltiples derivaciones ECG con la misma base de datos que utilizamos aquí, en su mayoría basados en aprendizaje profundo o *deep learning* por su traducción del inglés. Como se observa en la tabla comparativa en [Reyna 2021], otros autores obtuvieron puntuaciones sustancialmente superiores a las presentadas en este trabajo, presentando enfoques diferentes a los utilizados aquí.

En [Xiaoyu 2021], se construyó un modelo basado en SE-ResNet incorporando la detección de picos como una tarea auxiliar autosupervisada. Sus puntuaciones de *CM* fueron 0,55, 0,58, 0,58, 0,57 y 0,57 utilizando las combinaciones de 12, 6, 4, 3 y 2 derivaciones en el conjunto oficial y oculto para *test*.

En [Wickramasinghe 2021], se entrenaron dos CNN separadas utilizando segmentos de ECG de 20 segundos y su transformada rápida de Fourier. Sus puntuaciones de *CM* fueron 0,55, 0,51, 0,56, 0,55 y 0,56 utilizando las combinaciones de 12, 6, 4, 3 y 2 derivaciones en el conjunto oficial y oculto para *test*.

En [Aublin 2021], se diseñó un sistema de votación, donde los segmentos de ECG de 10 segundos alimentaban una CNN con un elevado número de capas para cada derivación disponible. Sus puntuaciones de *CM* fueron 0,48, 0,47, 0,47, 0,47 y 0,46 utilizando las combinaciones de 12, 6, 4, 3 y 2 derivaciones en el conjunto oficial y oculto para *test*.

Finalmente, en el contexto del *Special Issue* titulado *Classification of Multilead ECGs* de la revista *Physiological Measurement*, otros nueve autores mejoraron sus propuestas iniciales, validando sus modelos en los mismos conjuntos de datos ocultos. La Tabla 5.13 compara los resultados de nuestro trabajo con los otros nueve que finalmente pasaron el filtro de la revista para su publicación [Reyna 2022]. Ahí, los mejores valores de la métrica *CM* fueron de 0,62, 0,61 y 0,59 para los conjuntos de 12, 3 y 2 derivaciones respectivamente. De los diez trabajos publicados, el nuestro se sitúa en la séptima posición del *ranking*. Todos los modelos que mejoraron nuestro rendimiento fueron redes convolucionales de distintos tipos. Por otro lado, nuestra aproximación mejoró el rendimiento de tres modelos de clasificación que tenían los siguientes tipos de redes neuronales recurrentes como base de su sistema: *Convolutional Recurrent Neural Network* (CRNN), *Long Short-Term Memory* (LSTM), *Neural Basis Expansion Analysis for Interpretable Time Series* (N-BEATS) y *Gated Recurrent Unit* (GRU). En [Reyna 2022] también aparecen resultados de tres trabajos más (dos de ellos con mejores puntuaciones que las

nuestras, y otro con rendimientos de clasificación más bajos), que finalmente no fueron publicados.

Tabla 5.13. Comparación de nuestros resultados de clasificación con otros trabajos, utilizando la *Challenge Metric (CM)*, en el conjunto oficial de *test oculto*.

Estas puntuaciones fueron proporcionadas *online* en el contexto del *Special Issue* titulado *Classification of Multilead ECGs* de la revista *Physiological Measurement* [Reyna 2022].

Referencia	Modelo Utilizado	(CM)	(CM)	(CM)
		12 deriv.	3 deriv.	2 deriv.
Antoni 2022	CNN, ResNet50	0,62	0,61	0,59
Nejedly 2022	CNN, ResNet	0,59	0,60	0,59
Xu 2022	CNN SE_ResNet	0,57	0,58	0,57
Xia 2022	2D CNN	0,52	0,51	0,50
Aublin 2022	CNNs + FFNNs	0,48	0,46	0,45
Vázquez 2022	1D CNN	0,48	0,36	0,50
Jiménez 2022	FFNN, NB, k-medias	0,39	0,38	0,37
Kang 2022	CRNNs	0,38	0,40	0,35
Sawant 2022	LSTM	0,36	0,32	0,38
Puszkarski 2022	LSTM, N-BEATS, GRU	-0,12	-0,12	-0,08

Los trabajos mencionados anteriormente tuvieron resultados prometedores, pero ninguno de ellos evaluó su rendimiento con una única derivación ECG en la base de datos, tal y como hicimos nosotros. Además, nuestro conjunto de características extraídas podría ser evaluado de forma más sencilla por médicos expertos con el fin de comprender mejor los resultados de nuestros modelos. Finalmente, también destacamos que nuestro sistema fue diseñado de manera modular, lo que permite futuras actualizaciones y mejoras, como agregar nuevas clases, características y modelos de clasificación al sistema de entrenamiento y validación de manera sencilla. Por otro lado, nuestro sistema también permite reducir fácilmente el número de clases a detectar, algo que, aunque parezca contraintuitivo, lo haría más fiable al dar respuesta únicamente respecto de las afecciones cuyo coste-efectividad fuese equilibrado teniendo en cuenta los resultados de clasificación obtenidos.

5.5.4. Limitaciones y trabajos futuros

Este estudio presenta algunas limitaciones a tener en cuenta. A continuación, enumeramos algunas de ellas, sugiriendo ciertas líneas de trabajos futuros al respecto.

En primer lugar, sería interesante evaluar el rendimiento del detector de ondas R y T empleado, mejorando su rendimiento en aquellas circunstancias donde se requiera. La detección de estos puntos fiduciales es crucial, estando directamente relacionada con los rendimientos de clasificación, por lo que en un futuro sería interesante ampliar el estudio en este sentido.

El segundo punto a tratar sería la inclusión de características y/o reglas clínicas más específicas para cada una de las afecciones cardiacas a clasificar. Aunque este estudio contempla nueve categorías de características diferentes, es posible incluir otras que resultarían de gran ayuda en la clasificación, como la presencia de latidos ectópicos o análisis espectrales más elaborados. No obstante, esto iría en detrimento del coste computacional, lo que, por otro lado, hace que nuestra solución mantenga un buen compromiso para su implementación en dispositivos portátiles. Pero en circunstancias donde el coste computacional no sea relevante, la inclusión de tantas características útiles como sea posible es clave para mejorar los rendimientos de clasificación.

Finalmente, la inclusión de otros tipos de clasificadores, como SVM, árboles de decisión, sistemas de reglas, o redes convolucionales, puede ser decisivo a la hora de aumentar el rendimiento de clasificación. Una vez más, dependiendo de las circunstancias y el dispositivo donde se vayan a utilizar estos modelos, el coste computacional de inferencia debe ser tenido en cuenta.

En resumen, nuestro trabajo presenta todavía algunas imprecisiones en el diagnóstico de afecciones cardiacas. Por lo tanto, a pesar de los beneficios de la monitorización automática del ECG mediante dispositivos portátiles e inteligentes, se debe tener especial cuidado al presentar diagnósticos al usuario en un escenario no clínico. Por un lado, los falsos negativos podrían dar lugar a una falsa sensación de seguridad. Por otro lado, los falsos positivos podrían desencadenar una visita innecesaria al médico que, a gran escala, podría llevar al colapso de los centros de salud de manera absurda. La optimización futura de estos clasificadores podría evitar estos inconvenientes y circunstancias no deseadas.

5.5.5. Conclusiones

En este trabajo se ha presentado y evaluado una metodología para la detección de múltiples afecciones cardiacas a través de registros ECG que combina la extracción y selección de características, unidas a una aproximación de clasificación *One-vs-Rest* que utiliza FFNN, NB y un sistema híbrido como clasificadores binarios. Es interesante destacar que después de un análisis sistemático, los resultados de clasificación que utilizan sólo una o dos derivaciones ECG no estaban muy alejados de aquellos que usaron 12 derivaciones, mostrando costes computacionales más bajos y siendo más adecuados para dispositivos portátiles. Además, para algunas enfermedades cardiacas en concreto, el uso de una o dos derivaciones ECG mostró rendimientos de clasificación iguales o superiores que

otras configuraciones con más derivaciones. Mejorar la identificación de algunos ritmos cardíacos mediante la incorporación de características más específicas o reglas clínicas para aquellos casos en los que el rendimiento fue bajo es una dirección interesante a explorar en futuros trabajos.

Capítulo 6

Marcadores y clasificadores para detectar Miocardiopatía Arritmogénica

6.1. Introducción

La Miocardiopatía Arritmogénica (MCA) es una enfermedad cardiaca hereditaria que causa infiltraciones fibroadiposas en el tejido miocárdico [Marcus 1982], produciendo alteraciones de la conducción eléctrica que, finalmente, pueden desencadenar arritmias ventriculares. Esto provoca que la MCA sea una de las principales causas de muerte súbita en personas jóvenes por insuficiencia cardiaca, especialmente las asociadas al deporte [Suárez 2013]. Además, al tener una transmisión normalmente autosómica dominante, las posibilidades de padecer la enfermedad aumentan drásticamente entre los familiares directos de los afectados [Sharma 2022, Daliento 1995, Nava 1988]. La prevalencia de la MCA es de 1:2000-5000, considerándose una enfermedad rara de base genética, con un diagnóstico clínico complicado [Corrado 2020, Calkins 2017, McKenna 2017, Corrado 2011, Sen 2008].

Antes de 2020, dicho diagnóstico clínico se basaba en los criterios *Task Force* de 2010 (CTF) [Marcus 2010], especialmente dirigidos a la detección de las formas con afectación del ventrículo derecho (VD). Por otro lado, en los años posteriores se reconocieron formas dentro del espectro fenotípico de la MCA, cuya afectación predominante era en el ventrículo izquierdo (VI) y también biventricular (biV). No obstante, los CTF eran poco sensibles en el diagnóstico de estas formas predominantes VI y biV, al estar descritos principalmente para la forma VD.

En 2020 se revisaron los criterios diagnósticos de la MCA llegando al consenso *Padua Criteria* [Corrado 2020], basado principalmente en los CTF anteriores. Aquí se incorporaron nuevos criterios para el diagnóstico del fenotipo VI, en particular

mediante la incorporación de los hallazgos de caracterización tisular de la resonancia magnética cardiaca, así como algunas anomalías del ECG. No obstante, estos criterios para afectación del VI, en especial los del ECG, son todavía objeto de estudio y de mejora, debiendo ser todavía validados mediante estudios clínicos en cohortes más grandes de pacientes [Corrado 2021]. Por otro lado, no está descrito ningún criterio diagnóstico específico para la afectación biV ni en los CTF ni en el *Padua Criteria*.

Entre los criterios diagnósticos actuales se incluyen observaciones en la ecocardiografía, resonancia magnética nuclear, estudios de prevalencia entre familiares, análisis genético, histológico y del ECG de superficie. Estos criterios se agrupan en *mayores* o *menores* según su sensibilidad, siendo motivo de diagnóstico de MCA la presencia de dos criterios mayores, o uno mayor más dos menores, o cuatro criterios menores de diferentes categorías [Corrado 2020]. De entre todas las pruebas, el ECG estándar de 12 derivaciones es la técnica más simple y utilizada para detectar arritmias cardiacas. En este sentido, se han identificado ciertos cambios morfológicos en el ECG como indicativos de una posible MCA, mostrando alteraciones de la repolarización y la despolarización asociadas con la enfermedad. Entre otros criterios, se han descrito inversiones de la onda T en las derivaciones precordiales derechas (V1-V3), al igual que parámetros relacionados con la duración, amplitud y fragmentación del complejo QRS junto con la aparición de la onda épsilon en diferentes derivaciones [Pilichou 2016, Quarta 2012, Marcus 2010].

A pesar de lo mencionado anteriormente, otros indicadores del ECG que han demostrado eficacia diagnóstica en distintas patologías de origen ventricular no han sido validados en MCA.

Una de las técnicas a explorar es el Vectocardiograma (VCG), el cual es un método para analizar la trayectoria descrita por la actividad eléctrica cardiaca, obtenido a partir de las 12 derivaciones ECG estándar mediante la Transformada Inversa de *Dower* [Dower 1984, Edenbrandt 1988]. Se ha utilizado previamente para caracterizar distintas patologías, como por ejemplo el *flutter* auricular [Castells 2011], la hipertrofia ventricular izquierda y derecha [Titomir 2007], o estudiar recientemente las alteraciones cardiacas en población anciana [Bortolan 2022].

Por otro lado, el Análisis de Componentes Principales (PCA) también se ha utilizado con éxito en el procesamiento de señales de ECG con diferentes aplicaciones [Bortolan 2022, Feeny 2020, Bonizzi 2010, Castells 2007, Castells 2005a].

Finalmente, otra técnica de reconocimiento de patrones que puede aplicarse en la segmentación y clasificación de señales son los Modelos Ocultos de Markov, más conocidos como HMM por sus siglas en inglés (*Hidden Markov Models*). Estudios previos han demostrado su eficacia, en la detección de anomalías en el complejo P-QRS-T [Liang 2014], o en la clasificación del infarto de miocardio en registros ECG [Chang 2012], entre otras patologías. No obstante, estos modelos no han sido probados en la detección de la MCA.

Adicionalmente, otros trabajos han descrito diferencias significativas entre mujeres y hombres en distintos parámetros ECG, comenzando éstas en la adolescencia y evolucionando a partir de entonces [Pujadas 2022, Moss 2010, Macfarlane 1994]. Por lo tanto, también es necesario explorar la estratificación por sexos en la caracterización e identificación de la MCA.

Todo lo descrito anteriormente demuestra la necesidad de encontrar y validar criterios, indicadores y métodos que permitan mejorar el diagnóstico clínico y la detección de la MCA en sus formas de afectación VI y biV, utilizando registros ECG. Para ello, es posible explorar los criterios ya existentes para VD, alguna variación de éstos o los definidos recientemente para VI en el *Padua Criteria*, así como las distintas técnicas que hemos propuesto.

Por simplicidad, de aquí en adelante, cuando hablemos de MCA se entenderán incluidas únicamente las formas de afectación VI y biV, puesto que nuestro estudio pretende diferenciarse de los criterios ya existentes para VD.

6.1.1. Objetivos

Los objetivos detallados de este estudio son los siguientes:

- Analizar exhaustivamente registros cortos de ECG de 12 derivaciones de pacientes MCA, comparándolos con los de familiares no afectados de similar edad y sexo.
- Identificar y evaluar nuevos criterios basados en la amplitud, área y morfología de las distintas ondas del ECG que muestren rasgos característicos en pacientes de MCA.
- Evaluar la trayectoria del VCG, tanto de todo el ECG como la correspondiente al complejo QRS, proponiendo un conjunto de indicadores que muestre diferencias entre sujetos afectados por MCA y el grupo control procedente de cribados familiares.
- Proponer y evaluar nuevos criterios basados en los *eigenvalores* obtenidos del PCA correspondientes a registros cortos de ECG de 12 derivaciones, que muestren diferencias entre los sujetos afectados por MCA y el grupo control procedente de cribados familiares.
- Analizar las diferencias existentes en los criterios anteriores al estratificar las muestras por sexo.

- Evaluar distintos modelos de clasificación basados en regresión logística, que identifiquen a afectados y no afectados por MCA utilizando registros cortos de ECG de 12 derivaciones, contemplando también la estratificación por sexo de los pacientes.
- Evaluar los modelos ocultos de Markov en la clasificación de sujetos afectados por MCA y familiares que no padecen la enfermedad a través de registros cortos de ECG de 12 derivaciones.

6.2. Materiales

La base de datos de este trabajo se obtuvo en el transcurso de un estudio clínico en colaboración con la *Unidad de Cardiopatías Familiares del Hospital Universitario y Politécnico La Fe de València*. Los registros de ECG anonimizados y datos poblacionales utilizados se obtuvieron en el transcurso de tres años y medio.

El grupo de casos lo componen registros de 41 pacientes diagnosticados previamente con MCA (25 con afectación preferente del VI, 16 con afectación biV). Por otro lado, como grupo control se utilizaron registros de 46 sujetos que habían sido sometidos a pruebas diagnósticas de MCA debido a tener familiares directos afectados, pero con resultado negativo en el diagnóstico de la enfermedad, y sin hallazgo de la mutación genética. Es decir, el grupo control corresponde a personas negativas en las pruebas diagnósticas de MCA que participaron en un cribado intra-familiar.

Cada uno de estos registros se compone de señales en las 12 derivaciones ECG estándar, de una duración de 10 segundos y una frecuencia de muestreo de 500 Hz, adquiridas mediante un equipo *Philips PageWriter Touch*. La Tabla 6.1 muestra los datos poblacionales de la base de datos empleada. Respecto a las proporciones de sexo, las mujeres estaban ligeramente sobrerrepresentadas en el grupo control con un 61% de las muestras, mientras que en el grupo de afectados por MCA el 56% de las muestras correspondía a hombres. En esta tabla también aparece la edad media y su desviación estándar para cada grupo, no existiendo grandes diferencias entre los afectados por MCA y el grupo control, tal como también se aprecia en la Figura 6.1.

Tabla 6.1. Datos poblacionales de la base de datos de MCA empleada.

	Pacientes MCA (n=41)	Grupo Control (n=46)
Sexo Femenino	18 (43.9%)	28 (60.9%)
Sexo Masculino	23 (56.1%)	18 (39.1%)
Edad (años)	43,0±17,3	43,0±18,2

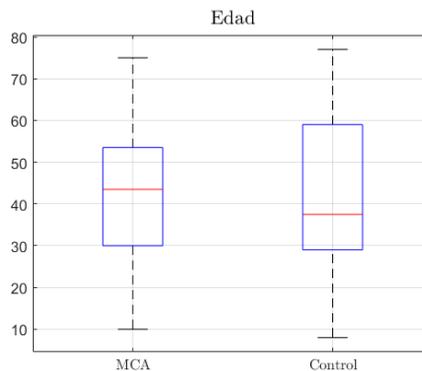


Figura 6.1. Edad de los participantes del estudio de MCA (diagrama de cajas).

6.3. Métodos

6.3.1. Preprocesado de la señal

Todos los registros ECG de la base de datos utilizada presentaban una duración de 10 segundos y una frecuencia de muestreo de 500 Hz, por lo que no fue necesario emplear ningún remuestreo. A cada uno de estos registros se le aplicó el preprocesado de la señal descrito a continuación.

En primer lugar, eliminamos los tramos con artefactos aberrantes en los que aparecían valores atípicos, causados principalmente por movimientos bruscos y repentinos de los pacientes, reflejados como cambios abruptos en el nivel de la señal. Para detectarlos, usamos una ventana deslizante de 0,5 segundos y calculamos sus valores máximo y mínimo. Si la diferencia estadística entre ventanas colindantes excedía los umbrales de entre cuatro y cinco desviaciones típicas respecto a la mediana, considerábamos anómalo dicho segmento y todos los valores de la señal dentro de esa ventana se establecían en cero.

Seguidamente se aplicó un filtro paso bajo de *Butterworth* de cuarto orden y 70 Hz para eliminar el ruido de alta frecuencia. El siguiente filtro utilizado fue de tipo *notch* IIR (*Infinite Impulse Response*) de segundo orden para 50 Hz con el fin de eliminar las interferencias de la red eléctrica. Finalmente, se aplicó el último filtro paso alto de 0,7 Hz, también de tipo *Butterworth* de cuarto orden, para eliminar los artefactos de baja frecuencia, así como la desviación de la línea base provocada principalmente por la respiración del paciente.

A continuación, eliminamos 0,5 segundos de señal al principio y al final de las 12 derivaciones de cada registro ECG, omitiendo de este modo la etapa de estabilización de los filtros. Esto nos condujo a que la duración de todas las muestras de la base de datos finalmente fuera de 9 segundos, no existiendo por lo tanto ninguna diferencia en el tiempo ni en el muestreo de todas ellas.

Como último paso de este preprocesado, restamos la moda, (*i.e.* el valor más frecuente) a cada una de las derivaciones ECG para centrar cada señal en cero. Para calcular la moda, se aplicó previamente un redondeo de tres decimales a los valores de las muestras, obteniendo así una moda más fiable para las señales empleadas, las cuales presentaban mucha variabilidad en sus últimos decimales.

Las Figura 6.2, 6.3 y 6.4 muestran el efecto de este preprocesado de señal en tres muestras diferentes de la base de datos, correspondientes cada un individuo del grupo control o afectados por MCA (biV y VI) respectivamente.

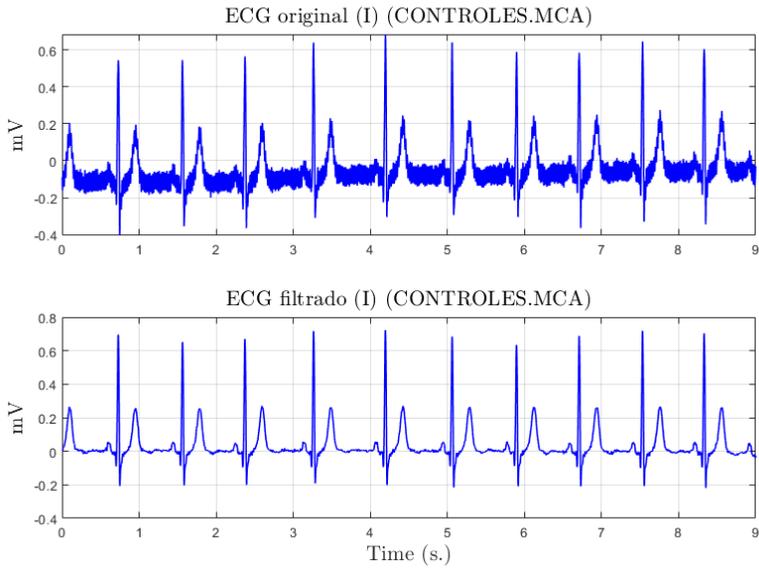


Figura 6.2. Filtrado de señal ECG de una persona del grupo control (derivación I).

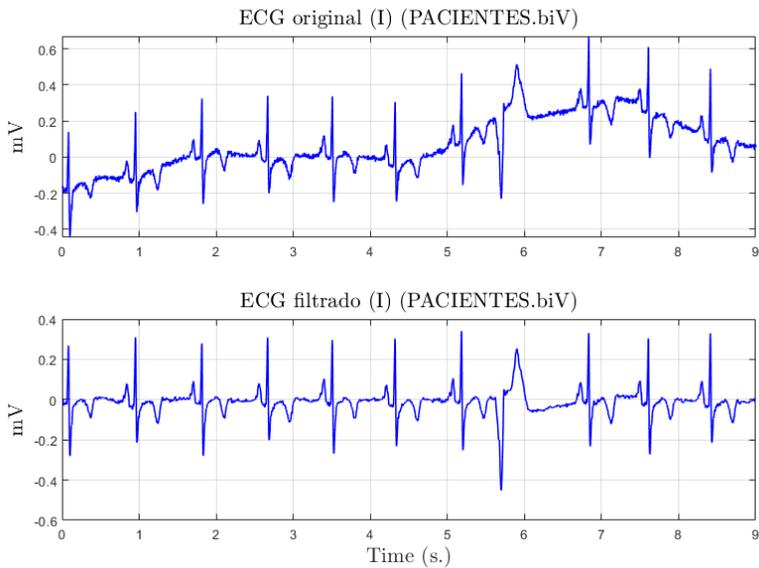


Figura 6.3. Filtrado de señal ECG de una persona afectada por MCA-biV (derivación I).

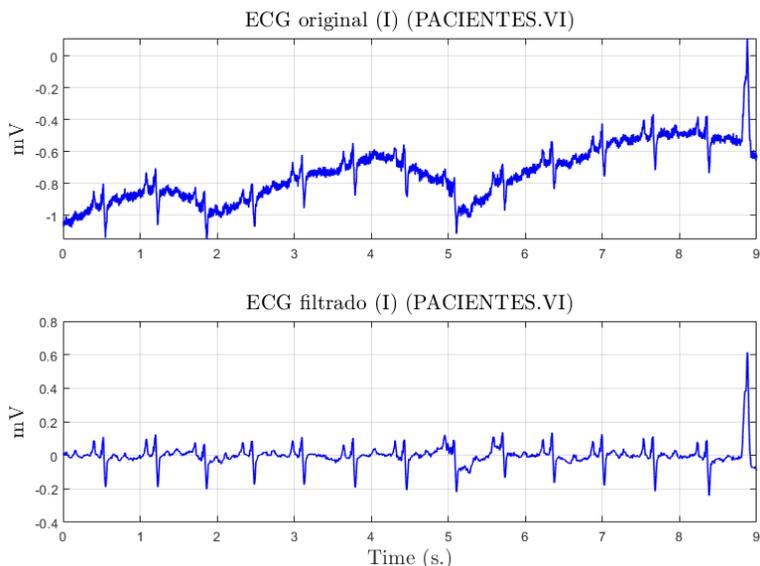


Figura 6.4. Filtrado de señal ECG de una persona afectada por MCA-VI (derivación I).

6.3.2. Extracción de parámetros basados en el ECG

A continuación, se describen los parámetros propuestos para caracterizar las señales ECG de 12 derivaciones estándar basadas principalmente en la amplitud y morfología del complejo QRS-T.

En primer lugar, se identificaron los picos de la onda R en todas las derivaciones mediante nuestro detector basado en el algoritmo de Pan y Tompkins y la primera derivada del ECG [Pan 1985]. Para evitar falsos positivos, se filtraron las detecciones de la onda R que no superasen un valor medio de 0,75 en el coeficiente de correlación con todas las demás detecciones, utilizando una ventana de 200 ms, eliminando así las detecciones en latidos ectópicos y artefactos.

Utilizando las marcas del detector anterior, se obtuvo la amplitud media de la onda R en cada derivación (*RAmp*).

También obtuvimos un valor booleano en cada derivación ECG que indicaba si la onda R sufría o no una fragmentación (*Rfrag*). Para ello, comprobamos si existían más de un cambio de signo en la segunda derivada de la señal correspondiente a los valores próximos del pico de la onda R. Si esta condición se cumplía en más de la mitad de las ondas R detectadas, el valor de *Rfrag* se establecía en verdadero. Para el caso contrario, donde el único cambio de pendiente correspondía, como era de esperar, al pico de la onda R, el valor de *Rfrag* se establecía en falso. Para cada registro ECG, se contabilizó el número de derivaciones con fragmentación en la onda R, parámetro al que denominamos *Num_Rfrag*.

Seguidamente se creó un patrón del complejo QRS en las 12 derivaciones estándar del ECG mediante promediado, utilizando una ventana de 200 ms centrada en las detecciones de la onda R de la etapa anterior. Del patrón de cada derivación ECG se extrajo su área utilizando integración numérica trapezoidal mediante la función *trapz* de Matlab. El valor de esta área es etiquetado aquí como *QRSArea*.

Se extrajo automáticamente dos características más del voltaje de la señal ECG para cada derivación. La primera fue el nivel medio de referencia (*midLevel*) para formas de onda de dos niveles (superior e inferior). La segunda fue la diferencia de amplitud entre los dos niveles (*difLevel*). Para ello, utilizamos las funciones *midcross* y *statelevels* de Matlab respectivamente.

Por otro lado, teniendo en cuenta un desplazamiento posterior al complejo QRS de 100 ms, definimos ahí una ventana de 300 ms como la zona correspondiente a la onda T. Los valores de la onda T en las derivaciones I, II, V4, V5 y V6 normalmente son positivos y su inversión puede tener valor diagnóstico. Es por ello que, para este conjunto de derivaciones obtuvimos un valor booleano que indicaba si la onda T era o no negativa (*TNeg*), así como el número total de ondas T que cumplían dicha condición (*Num_TNeg*). Además, también obtuvimos para este grupo de derivaciones la media del área de la onda T (*TArea*) utilizando una vez más la integración numérica trapezoidal en la ventana de 300 ms definida anteriormente.

Finalmente, obtuvimos el valor del parámetro *TWAD* (*T-Wave Area Dispersion*) descrito en [Simov 2016] y [Bortolan 2022]. Este parámetro se calculó como la media de las *TAreas* normalizadas por el valor absoluto del área máxima, tal como se indica en la siguiente expresión

$$TWAD = \frac{1}{N} \sum_{i=1}^N \frac{TArea_i}{\max(|TArea_{i:N}|)} \mid i \in \{I, II, V_4, V_5, V_6\} \quad (6.1)$$

El rango de valores de *TWAD* se encuentra entre -1 (todas las áreas negativas idénticas) y +1 (áreas positivas idénticas).

En resumen, se extrajeron un total de 73 parámetros de cada registro ECG relacionados con la amplitud y morfología de la señal. De las 12 derivaciones estándar, se obtuvieron los valores de *RAmp*, *Rfrag*, *QRSArea*, *midLevel* y *difLevel*. De las derivaciones I, II, V4, V5 y V6 se obtuvieron los valores de *TArea* y *TNeg*. Finalmente, como parámetros individuales que agrupaban información de distintas derivaciones extrajimos los 3 indicadores denominados *Num_RFrag*, *Num_TNeg* y *TWAD*.

6.3.3. Extracción de parámetros basados en el Vectocardiograma

A continuación, se describen los métodos propuestos para extraer y caracterizar el VCG a partir de registros ECG de 12 derivaciones estándar. Además de caracterizar la trayectoria descrita por el VCG correspondiente a los 9 segundos de señal de cada registro ECG filtrado, también se pretende analizar la trayectoria descrita por el complejo QRS en el VCG.

Para obtener el patrón del lazo que describe el VCG durante el complejo QRS, en primer lugar, se identificaron los picos de la onda R en la derivación V5 mediante el detector basado en el algoritmo de Pan y Tompkins [Pan 1985] utilizado en el apartado anterior. Seguidamente se creó un patrón del complejo QRS en las 12 derivaciones estándar del ECG mediante promediado, utilizando una ventana de 200 ms centrada en las detecciones de la onda R de la derivación V5 con el fin de sincronizar los patrones de todas las derivaciones.

Para la obtención del vectocardiograma se aplicó la *Transformada Inversa de Dower* [Dower 1984] tanto sobre toda la señal ECG como sobre el patrón QRS obteniendo así dos representaciones en el espacio tridimensional (derivaciones X, Y, Z) a las que denominamos *VCG* y *VCG(QRS)*.

A continuación, se obtuvieron los parámetros de la regresión lineal múltiple que definen el plano que mejor se ajustaba a cada VCG utilizando la función *regress* de Matlab. Para medir el grado de torsión del VCG en el espacio 3D se utilizaron las 5 métricas descritas a continuación.

La primera métrica consistió en el estadístico R^2 obtenido durante el proceso de ajuste del plano, donde un valor de 1 indica un ajuste perfecto, actuando como límite superior.

Para las dos siguientes métricas se utilizó la distancia euclídea D entre el vector VCG y el vector \hat{Y} resultante de proyectar dicho VCG sobre el plano ajustado anteriormente.

$$D = \sqrt{\sum (VCG - \hat{Y})^2} \quad (6.2)$$

Para obtener medidas comparables entre sujetos independientemente del número N de puntos que contiene cada VCG, se utilizaron los valores $D1$ y $D2$, definidos como

$$D1 = \frac{D}{N} \quad (6.3)$$

$$D2 = (\sqrt[N]{D} - 1) \times 1.0e^3 \quad (6.4)$$

$D1$ puede describirse como la media de la distancia euclídea al mejor plano de ajuste, mientras que $D2$ es una aproximación de la distancia geométrica al mejor plano de ajuste donde se intenta que el rango de valores no se aleje demasiado del 1 en caso de distancias grandes en los registros utilizados.

La cuarta métrica utilizada en este estudio intenta capturar las variaciones locales con el fin de evaluar la complejidad de la trayectoria. Cuantos más cambios de dirección, más complejo es el camino que forma el lazo VCG. Por lo tanto, la complejidad puede considerarse como la acumulación de variaciones instantáneas del ángulo θ_i , donde cada i -ésimo punto se calcula de la siguiente manera

$$\theta_i = \cos^{-1} \left(\frac{(x_i - x_{i-1})^T (x_{i+1} - x_i)}{\|x_i - x_{i-1}\| \|x_{i+1} - x_i\|} \right) \quad (6.5)$$

La acumulación de estas variaciones angulares entre los puntos x_i durante toda la trayectoria del VCG tiene como cota inferior $2\pi rad$, valor obtenido en el caso más sencillo; es decir, si el vector describe un lazo cerrado, similar a una elipse dibujada en un plano 2D. Al no existir cota superior para este parámetro, podemos definir su simplicidad (S) como

$$S = \frac{2\pi}{\sum_i \theta_i} \in [0, 1] \quad (6.6)$$

La cota inferior de S indica una trayectoria muy compleja, mientras que la cota superior muestra una trayectoria sin variaciones locales [Castells 2011].

Por otro lado, en las derivaciones X-Y-Z del VCG, la magnitud de los vectores V que contienen n muestras puede ser calculada mediante la siguiente expresión

$$V_n = \sqrt{X_n^2 + Y_n^2 + Z_n^2} \quad (6.7)$$

La quinta y última métrica utilizada es el ratio (RT) del vector V_n definido como

$$RT = \frac{\max(V_n)}{\text{media}(V_n)} \quad (6.8)$$

Este parámetro RT fue definido en [Kallert 1999], y utilizado para la caracterización de la morfología del VCG en diversas afecciones cardiovasculares [Bortolan 2001, Bortolan 2002].

En conclusión, extrajimos los siguientes 5 parámetros tanto del VCG, como del VCG(QRS): el estadístico de mejor ajuste al plano R2, las medidas de distancia a

este plano $D1$ y $D2$, la métrica de simplicidad S , y la ratio entre la magnitud máxima y media del vector al que denominamos RT . Esto sumó un total de 10 indicadores derivados de cada registro ECG.

6.3.4. Extracción de parámetros basados en el Análisis de Componentes Principales

El Análisis de Componentes Principales (PCA) es una técnica estadística multivariante que permite la reducción de la dimensionalidad de grandes conjuntos de datos correlacionados en otro conjunto de datos con un menor número de dimensiones que el inicial. Cada una de estas nuevas dimensiones se denomina *componente principal* y puede representar mejor la varianza de los datos originales [Jolliffe 2002].

A continuación, se describen los parámetros propuestos basados en la descomposición de los *eigenvalores* (también llamados valores propios) obtenidos a partir del PCA de los registros ECG de 12 derivaciones.

Una vez filtradas las señales, realizamos un PCA de cada registro ECG. Los datos de las 12 derivaciones se organizaron en una matriz de 12 filas y n columnas, siendo n el número de muestras de cada registro. A continuación, realizamos una descomposición de valores singulares de la matriz de datos. Este PCA transforma el ECG en parámetros virtuales que son mutuamente independientes (es decir, ortogonales), donde los 3 primeros componentes principales contienen toda la información que representa el concepto vectorial de un dipolo eléctrico. Por otro lado, los componentes principales restantes representan los componentes no dipolares en el ECG. Para cuantificar la varianza en componentes dipolares y no dipolares, extrajimos los primeros 8 *eigenvalores*.

Cada *eigenvalor* representa la varianza de cada componente principal correspondiente a un registro ECG. La suma de los 8 *eigenvalores* representa el 100% de la varianza en los componentes principales. El primer *eigenvalor* contiene la varianza más alta ya que corresponde al primer y más importante componente principal. Por otro lado, el octavo *eigenvalor* contiene la varianza más baja. A partir de las definiciones anteriores, propusimos los dos criterios siguientes para ayudar al diagnóstico de la MCA.

El primer criterio ($C1$) representa el porcentaje de varianza contenida en *eigenvalores* de componentes principales no dipolares y se define como

$$C1 = \frac{\sum_{i=4}^8 \text{eigenvalor}_i}{\sum_{i=1}^8 \text{eigenvalor}_i} \quad (6.9)$$

Es decir, la suma de la varianza del *eigenvalor* cuarto al octavo dividida por la varianza total contenida en los ocho componentes. Esto nos da una medida de energía de la varianza contenida en los componentes no dipolares. Los valores de

este criterio se definen en el intervalo (0, 1), mostrando el límite inferior varianza nula en los componentes no dipolares.

El segundo criterio propuesto ($C2$) representa el nivel de energía contenido en el tercer *eigenvalor* entre los componentes dipolares (los tres primeros) y se define como

$$C2 = \frac{\text{eigenvalor}_3}{\text{eigenvalor}_1 + \text{eigenvalor}_2} \quad (6.10)$$

Esta métrica explica la variación de energía entre los componentes dipolares. Los valores de $C2$ también se definen en el intervalo (0, 1), mostrando varianza nula en el límite inferior para el tercer componente dipolar. $C2$ indica el nivel de contribución de la activación del dipolo en la tercera dimensión. En caso de un valor muy bajo, la activación podría explicarse desde un espacio bidimensional, es decir, en un plano.

Al igual que en los parámetros extraídos a partir del VCG, estos criterios basados en PCA se obtuvieron utilizando por un lado toda la señal del registro ECG, y por otro únicamente la señal que formaba el patrón QRS en las 12 derivaciones. De este modo, extrajimos de cada registro ECG un total de 4 valores etiquetados con el sufijo PCA y PCA(QRS) indistintamente.

6.3.5. Resumen de parámetros y análisis estadístico

La Tabla 6.2 muestra un resumen de las características extraídas de cada registro ECG, donde 73 correspondían a parámetros basados en la amplitud y morfología de las ondas QRS y T, 10 fueron extraídos a partir del VCG y finalmente, 4 se derivaron del análisis PCA. En total, el conjunto de datos constaba de 87 características extraídas de cada una de las 87 muestras de nuestra base de datos.

Una vez obtenidos los valores de los parámetros enumerados en la Tabla 6.2 para cada registro ECG, en el caso de las variables numéricas continuas, utilizamos la prueba de *Kolmogorov-Smirnov* para corroborar la normalidad en los datos. Seguidamente, comparamos los resultados en ambos grupos de sujetos con el fin de encontrar diferencias estadísticamente significativas mediante un test *t de Student*.

Por otro lado, para el caso de las variables binarias, empleamos el test χ^2 de *Pearson* con el mismo objetivo de hallar diferencias significativas entre el grupo de casos y el de controles.

En ambas pruebas (*t-test* y χ^2), consideramos un valor $p < 0,01$ como estadísticamente significativo, informando únicamente los resultados de aquellos parámetros que superaron este criterio.

Los valores de los datos numéricos se reportan mediante la media \pm desviación estándar, mientras que en el caso de los valores binarios se muestra el número de

muestras positivas, así como su porcentaje respecto del total del grupo correspondiente.

Tabla 6.2. Resumen de las características extraídas de cada registro ECG.

Características	Derivaciones	Descripción
<i>RAmp</i>	12	Amplitud media de la onda R
<i>Rfrag</i>	12	Onda R fragmentada
<i>Num_Rfrag</i>	-	Nº total de ondas R fragmentadas
<i>QRSArea</i>	12	Área del complejo QRS
<i>midLevel</i>	12	Amplitud media para señales de dos niveles
<i>difLevel</i>	12	Diferencia de niveles de amplitud
<i>TArea</i>	<i>I, II, V₄, V₅, V₆</i>	Área de la onda T
<i>TWAD</i>	-	Medida de dispersión del área en ondas T.
<i>TNeg</i>	<i>I, II, V₄, V₅, V₆</i>	Onda T negativa
<i>Num_TNeg</i>	-	Nº total de ondas T negativas detectadas
<i>R2(VCG)</i>	-	Estadístico R2 de ajuste al plano
<i>D1(VCG)</i>	-	Distancia euclídea media al plano 3D
<i>D2(VCG)</i>	-	Distancia geométrica media al plano 3D
<i>S(VCG)</i>	-	Medida de torsión del VCG
<i>RT(VCG)</i>	-	Ratio de la magnitud máx. y media del VCG
<i>R2(VCG(QRS))</i>	-	Ídem R2(VCG) pero usando el patrón QRS
<i>D1(VCG(QRS))</i>	-	Ídem D1(VCG) pero usando el patrón QRS
<i>D2(VCG(QRS))</i>	-	Ídem D2(VCG) pero usando el patrón QRS
<i>S(VCG(QRS))</i>	-	Ídem S(VCG) pero usando el patrón QRS
<i>RT(VCG(QRS))</i>	-	Ídem RT(VCG) pero usando el patrón QRS
<i>C1(PCA)</i>	-	Varianza en eigenvalores no dipolares
<i>C2(PCA)</i>	-	Varianza 3er eigenvalor en comp. dipolares
<i>C1(PCA(QRS))</i>	-	Ídem C1(PCA) pero para el patrón QRS
<i>C2(PCA(QRS))</i>	-	Ídem C2(PCA) pero para el patrón QRS
Total: 87 características		

Además de obtener la significancia estadística utilizando todas las muestras de la base de datos para encontrar diferencias entre afectados por MCA y el grupo control, también se estratificó por sexos. De este modo, se dividió el conjunto de datos en dos subconjuntos: uno femenino y otro masculino. En ambos se realizó el mismo análisis estadístico que en el conjunto total de datos. Estos tres conjuntos de datos se identifican en este trabajo del siguiente modo: Conjunto de datos total (D_t), subconjunto de datos femenino (D_{fem}), subconjunto de datos masculino (D_{mas}).

6.3.6. Modelos de Regresión Logística y selección de características

Para cada conjunto de datos (D_t, D_{fem}, D_{mas}), se crearon varios modelos de clasificación mediante una regresión logística (RL) y un conjunto de características seleccionadas con el fin de detectar afectados por MCA.

Como métrica básica del rendimiento de clasificación se empleó la media geométrica entre sensibilidad y especificidad definida como

$$G = \sqrt{\text{Sensibilidad} \times \text{Especificidad}} \quad (6.11)$$

Para determinar el umbral (th) de la RL para la clasificación entre afectados por MCA y el grupo control se utilizó el valor que maximizaba la métrica G en el rango (0, 1) entre todas las salidas de las muestras de entrenamiento. Una vez determinado dicho umbral, la regla de clasificación que seguimos fue la siguiente:

$$\hat{c}_i = \begin{cases} 1 & \text{si } \hat{y}_i \geq th \\ 0 & \text{si } \hat{y}_i < th \end{cases} \quad (6.12)$$

donde \hat{y}_i representa la salida de la RL para una muestra i , mientras que \hat{c}_i representa la clase asignada, siendo 1 el grupo de afectados por MCA y 0 el grupo control.

Se empleó el método *Leave One Out Cross-Validation (LOOCV)* como estrategia de validación del rendimiento de clasificación de cada modelo RL. Este método iterativo emplea inicialmente como conjunto de entrenamiento todas las muestras disponibles excepto una, que se excluye para emplearla como validación. Este proceso se repite tantas veces como muestras disponibles existen, excluyendo en cada iteración una observación distinta. Finalmente, estimamos la métrica G del modelo a partir de las clasificaciones realizadas en todas las iteraciones con la muestra de validación, vector al que denominamos \hat{C} . Este proceso de validación requiere que el modelo sea reajustado y validado tantas veces como muestras existan en el conjunto de datos.

El conjunto de características a seleccionar en cada conjunto de datos se compuso inicial y exclusivamente por aquellas que habían mostrado significancia estadística con un valor de p menor que 0,01 en el correspondiente test estadístico.

A cada conjunto de datos, se aplicó los métodos de selección de características descritos a continuación, donde el vector $\delta = [\delta_1, \delta_2, \dots, \delta_n]$ representa el conjunto de características estadísticamente significativas, mientras que el vector $\lambda = [\lambda_1, \lambda_2, \dots, \lambda_m]$ representa el conjunto de características seleccionadas.

El primer método de selección de características empleado fue una variante del algoritmo *Forward-Selection (FS)* [Kohavi 1997]. En él, inicialmente el vector λ no

contiene ninguna característica. Cada iteración del algoritmo *FS* añade una a una todas las características del vector δ todavía no seleccionada al conjunto λ y obtiene el valor de G mediante la validación *LOOCV* para cada δ_i añadida. Si una vez finalizada la iteración, el mejor valor de G obtenido hasta el momento en anteriores iteraciones no aumenta, se finaliza el proceso. Por el contrario, si al añadir una característica δ_i al conjunto λ el valor de G mejora respecto al obtenido en la iteración anterior, se selecciona aquella que maximiza este valor G en la iteración actual. A continuación, se continúa con la siguiente iteración del algoritmo repitiendo este proceso hasta que se cumple la condición de parada o el conjunto λ crece hasta ser el mismo que el conjunto δ .

El segundo método de selección de características empleado fue una variante del algoritmo *Backward-Elimination (BE)* [Kohabi 1997], donde el conjunto λ contiene inicialmente todas las características estadísticamente significativas que aparecían en δ . Cada iteración del algoritmo elimina una a una las características del conjunto λ , obteniendo el valor de G mediante la validación *LOOCV* para cada λ_i eliminada. Al igual que en la aproximación *FS*, si al finalizar la iteración del algoritmo *BE*, la eliminación de algún parámetro no hacía mejorar el valor de G de la anterior iteración, se finalizaba el proceso de selección. No obstante, en este caso utilizamos una condición de parada más relajada, permitiendo continuar con la siguiente iteración si la eliminación de una λ_i presentaba una diferencia menor o igual a 0,02 respecto al valor de G de la iteración anterior, permitiendo a esta heurística obtener conjuntos λ más reducidos. Si dicha condición de parada no se cumplía, indicaba que eliminar alguna característica mejoraba o casi igualaba el valor de G de la iteración anterior, por lo que se seleccionó para su eliminación la λ_i que maximizaba el valor de G en cada iteración. A continuación, se continúa con la siguiente iteración del algoritmo repitiendo este proceso hasta que se cumple la condición de parada o sólo existe una característica en el conjunto λ .

Finalmente, para conjuntos de características δ reducidos, también se realizó una selección de características por fuerza bruta (*ZB*); es decir, se probaron todas las combinaciones posibles de características, seleccionando aquella que maximizaba el valor de G en la validación *leave-one-out*. Esto se pudo hacer sólo para aquellos conjuntos que no superaron 14 características, ya que la selección por fuerza bruta sufre una explosión combinatoria del orden de $n!$, intratable para valores de n superiores.

A estos conjuntos de características seleccionadas, los identificaremos por λ_{FS} , λ_{BE} , λ_{ZB} respectivamente según el método de selección empleado (*FS*, *BE*, *ZB*).

Para cada combinación viable entre el conjunto de datos (D_t , D_{fem} , D_{mas}) empleado y las características seleccionadas (λ_{FS} , λ_{BE} , λ_{ZB}), calculamos los valores de sensibilidad, especificidad y G utilizando el correspondiente vector \hat{C} . Además, también se calculó la precisión con su intervalo de confianza (IC) al 5% de

significancia estadística (α), evaluando con todas estas métricas el rendimiento de cada modelo de clasificación en la *LOOCV*.

Finalmente, definimos un último clasificador denominado D_t^* que se corresponde con la regla de clasificación siguiente:

$$\hat{c}_i = D_t^*(i) = \begin{cases} D_{fem}^* \lambda_x(i) & \text{si } i \in D_{fem} \\ D_{mas}^* \lambda_y(i) & \text{si } i \in D_{mas} \end{cases} \quad (6.13)$$

donde i representa la muestra a clasificar, \hat{c}_i la clase asignada a dicha muestra, mientras que $D_{fem}^* \lambda_x$ y $D_{mas}^* \lambda_y$ denotan las RL que mayor valor de G obtuvieron en su validación *LOOCV*, entre los distintos métodos de selección de características y subconjuntos de datos estratificados por sexo.

Resumiendo, generamos y evaluamos los nueve modelos de regresión logística siguientes:

- $D_t \lambda_{FS}$ Todos los datos, selección *FS*.
- $D_t \lambda_{BE}$ Todos los datos, selección *BE*.
- $D_{fem} \lambda_{FS}$ Subconjunto de datos sexo femenino, selección *FS*.
- $D_{fem} \lambda_{BE}$ Subconjunto de datos sexo femenino, selección *BE*.
- $D_{fem} \lambda_{ZB}$ Subconjunto de datos sexo femenino, selección *ZB*.
- $D_{mas} \lambda_{FS}$ Subconjunto de datos sexo masculino, selección *FS*.
- $D_{mas} \lambda_{BE}$ Subconjunto de datos sexo masculino, selección *BE*.
- $D_{mas} \lambda_{ZB}$ Subconjunto de datos sexo femenino, selección *ZB*.
- D_t^* Todos los datos, regla de decisión entre dos RL según sexo.

6.3.7. Generación del conjunto de datos para el entrenamiento de modelos ocultos de Markov

Para cada registro ECG de 12 derivaciones preprocesado, calculamos las derivadas primera y segunda. Como resultado, obtuvimos un total de 36 series temporales teniendo en cuenta las 12 señales originales.

Además, para cada uno de estos registros ECG preprocesados calculamos el VCG a través de la Transformada Inversa de *Dower* [Dower 1984], obteniendo así una representación tridimensional de la actividad eléctrica cardiaca. Este VCG agregó 3 series temporales más a nuestro conjunto de datos.

Finalmente, se realizó el mismo PCA utilizado en el anterior apartado 6.3.4. para cada registro ECG, agregando los 3 primeros componentes dipolares como series temporales a nuestro conjunto de datos. Es decir, transformamos las 12 derivaciones de los registros ECG en 3 series temporales mutuamente independientes correspondientes a las primeras 3 componentes principales obtenidas

mediante PCA. Utilizamos estos componentes del PCA ya que normalmente contienen la mayor parte de la información contenida en el ECG [Castells 2007].

Resumiendo todo lo anterior, de cada registro ECG creamos uno nuevo compuesto por 42 series temporales (12 derivaciones ECG y sus derivadas primera y segunda; el VCG compuesto por 3 vectores; finalmente, los 3 primeros componentes principales del PCA). Este nuevo conjunto de registros se utilizó como entrada en los procesos de entrenamiento y validación de modelos HMM para detectar la MCA.

6.3.8. División del conjunto de datos para entrenamiento y test de modelos ocultos de Markov

El ciclo de vida de un problema de reconocimiento de patrones puede dividirse en dos etapas principales: entrenamiento y reconocimiento. Durante la primera etapa se utiliza el conjunto de datos de entrenamiento para construir el modelo, que en este caso es un HMM. La segunda etapa utiliza las muestras restantes como conjunto de prueba, llamado también comúnmente conjunto de *test*, para evaluar el rendimiento del modelo durante la etapa de reconocimiento. Esta estrategia, llamada evaluación *hold-out*, se utilizó durante el presente estudio, donde se excluyó parte del conjunto de datos del entrenamiento con el fin de evaluar la precisión de los modelos con datos desconocidos por éstos.

La Tabla 6.3 muestra la cantidad de muestras asignadas a cada una de las dos divisiones de nuestro conjunto de datos original, según su pertenencia a los grupos de afectados por MCA o Control. En el caso de pacientes de MCA, se indica el número de afectados en el VI, y biV para cada caso. El 70% de los registros se usaron para entrenamiento y el 30% restantes para *test*. Dicha división se realizó al azar una vez y se usó durante los pasos de entrenamiento y *test* de cada HMM durante la experimentación. La prevalencia original de la MCA en nuestra base de datos era del 47,1%. Por otro lado, la división propuesta estableció una prevalencia del 47,5% en el conjunto de entrenamiento y del 46,1% en el conjunto de *test*, valores prácticamente idénticos a la prevalencia original de la base de datos.

Tabla 6.3. Registros ECG asignados a los conjuntos de entrenamiento y test para modelos HMM.

Registros ECG	MCA	Controles	Total
Entrenamiento	29 (18 VI, 11 biV)	32	61 (70%)
Test	12 (7 VI, 5 biV)	14	26 (30 %)
Total	41 (25 VI, 16 biV)	46	87

6.3.9. Entrenamiento de modelos ocultos de Markov con HTK

A continuación, se presenta la metodología de entrenamiento y validación de un conjunto de modelos HMM utilizando el kit de herramientas HTK (*Hidden Markov Model Toolkit*, versión 3.4.1) [Young 2006]. En este estudio evaluamos su precisión, sensibilidad y especificidad en la detección de sujetos afectados por MCA con afectación VI o biV.

Los HMM son modelos estadísticos basados en estados que pueden utilizarse para representar observaciones individuales de una secuencia. La idea general es que cada estado debe corresponder a una sección de la secuencia. Dentro del extenso marco teórico relativo a los modelos ocultos de Markov se han propuesto distintas soluciones a problemas de clasificación de series temporales descritas en la literatura, revisadas en [Khalifa 2021], y siendo el tutorial de *Rabiner* uno de los más conocidos [Rabiner 2013].

Para esta clasificación de series temporales, primero se utiliza el proceso iterativo del algoritmo de *Baum Welch* con el objetivo de estimar los parámetros del HMM durante la etapa de entrenamiento. Finalmente, se propone el algoritmo de *Viterbi* como una solución de decodificación para encontrar el siguiente estado del sistema más probable en función de su estado actual. Este proceso de transición entre estados es utilizado durante la etapa de reconocimiento de patrones [Abou 2015].

En relación con todo esto, HTK es un conjunto de herramientas *software* para construir y manipular modelos de tipo HMM utilizando muestras de entrenamiento para finalmente decodificar (clasificar en nuestro caso) otras muestras desconocidas. Además, HTK admite modelos HMM que utilizan mezcla de gaussianas, proporcionando herramientas para su entrenamiento, test y análisis de resultados. Aunque está destinado principalmente al reconocimiento de voz, también se ha utilizado en otras aplicaciones de reconocimiento de patrones en series temporales, como la segmentación de electromiogramas de afectados por la enfermedad de *Parkinson* [Bengacemi 2021], o la clasificación de fonocardiogramas [Touahria 2021].

La Figura 6.5 ilustra un esquema básico de un modelo HMM de topología *left-right*, utilizada en HTK, junto con sus parámetros más relevantes descritos a continuación:

- Número de estados ocultos (nE). Este parámetro del modelo es arbitrario y en este estudio se pretende ajustar el mejor valor posible para nuestros datos.
- Distribución del estado inicial ($\pi = [\pi_j]$): Vector cuyos elementos indican la probabilidad de que el HMM comience en cada estado j . En HTK este

parámetro no se utiliza ya que al usar una topología *left-right* el primer estado siempre corresponde al de más a la izquierda.

- Distribución de probabilidad de transición entre estados ($A = [a_{ij}]$): Matriz cuyos elementos representan la probabilidad de que el HMM transite del estado actual i al siguiente estado j .
- Distribuciones de probabilidad de emisión ($B = [b_j(O_t)]$): Colección de nE distribuciones de probabilidad continuas multivariantes (una por cada estado) que modelan la probabilidad de que el HMM genere una observación O en el instante de tiempo t , dado el estado actual j . Puesto que estamos considerando secuencias de observación multivariadas, se requiere una distribución de emisión multivariante. HTK utiliza una mezcla de distribuciones gaussianas a tal efecto descrita en [Young 2006].
- Número de gaussianas de la mezcla por estado (nG): Como se ha mencionado en el parámetro anterior, HTK modela las probabilidades de emisión $b_j(O_t)$ mediante una mezcla de gaussianas en cada uno de los estados del HMM. En este estudio, el número de gaussianas de la mezcla (nG) también se considera un parámetro del modelo a ajustar.

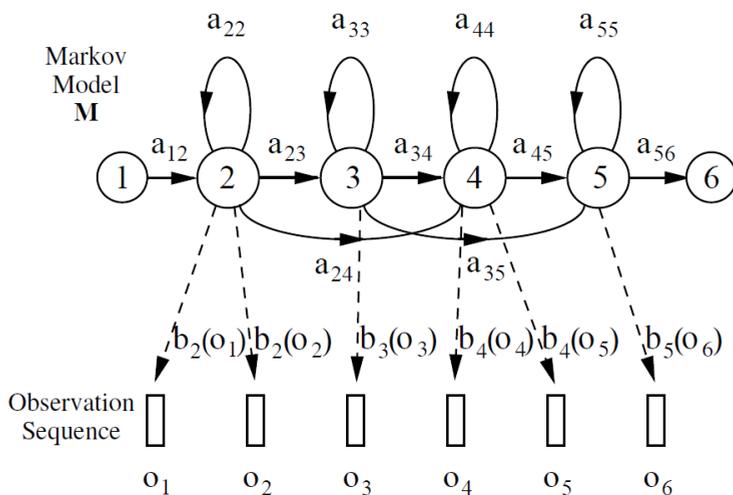


Figura 6.5. Esquema básico de un Modelo Oculto de Markov con topología *left-right* [Young 2006].

En este estudio llevamos a cabo una búsqueda de los mejores parámetros en modelos HMM combinando distintos valores en el número de estados (nE) y en el número de gaussianas utilizadas por estado (nG).

La búsqueda en el número de estados se realizó en el rango (20, 60) con pasos de 5, siguiendo la siguiente secuencia:

$$nE = \{20, 25, 30, \dots, 60\} \quad (6.14)$$

Por otro lado, la búsqueda en el número de gaussianas de las distribuciones de probabilidad de emisión por estado se realizó en el rango (1, 16) con pasos de potencias de 2 con el fin de permitir duplicar el número de gaussianas de cada estado. Utilizamos el comando HHed de HTK para *editar/modificar* este número de gaussianas por estado en determinadas iteraciones de nuestro proceso de aprendizaje. Es decir, nG tomó los siguientes valores en la búsqueda de parámetros:

$$nG = 2^{\{0,1,2,3,4\}} = \{1, 2, 4, 8, 16\} \quad (6.15)$$

De este modo, entrenamos un conjunto de 45 modelos HMM correspondientes a tuplas con la forma (nE, nG).

Todos los demás parámetros del modelo HMM se obtuvieron durante el proceso de entrenamiento mediante los algoritmos de *Baum-Welch* y *Viterbi*, utilizando las muestras del conjunto de entrenamiento previamente etiquetadas. En este sentido, el proceso de aprendizaje de *Baum-Welch* es una adaptación del algoritmo *Expectation-Maximization* a los HMM.

Denotamos el HMM que modela la clase c del siguiente modo:

$$\lambda_c = \left(A_c = [a_1, a_2, \dots, a_{nE}], B_c = \begin{bmatrix} b_{1,1} & \dots & b_{1,nG} \\ \vdots & \ddots & \vdots \\ b_{nE,1} & \dots & b_{nE,nG} \end{bmatrix} \right) \quad (6.16)$$

El ajuste de los parámetros A y B se realizó mediante el comando HERest de HTK, donde se aplica una iteración de reestimación de estos valores. La distribución del estado inicial π no aparece en el modelo al ser su estado inicial siempre el de más a la izquierda en la topología *left-right* utilizada.

Por otra parte, para una observación O (un registro ECG en nuestro caso), el algoritmo de *Viterbi* proporciona una estimación de la secuencia de estados del HMM más probable que haya generado dicha observación, denotada aquí por:

$$\widehat{X}_v = [x_1, x_2, \dots, x_T] \quad (6.17)$$

donde x_i es un estado concreto del modelo que puede aparecer una o más veces en la secuencia.

Con los parámetros del HMM ya establecidos por el algoritmo de *Baum-Welch*, y conocida la secuencia de estados más probable por *Viterbi* (\widehat{X}_v) para una observación O , la probabilidad de que el modelo haya generado dicha secuencia viene determinada por la siguiente expresión:

$$\hat{p}(O|\lambda_c) = a_{x(1)x(2)} \prod_{t=1}^T b_{x(t)}(O_t) a_{x(t)x(t+1)} \quad (6.18)$$

Teniendo todo lo anterior en cuenta, para cada uno de las tuplas (nE, nG) entrenamos dos modelos distintos mediante HTK que denotamos del siguiente modo:

$$\lambda_{mca} = (A_{mca}, B_{mca}) \quad (6.19)$$

$$\lambda_{control} = (A_{control}, B_{control}) \quad (6.20)$$

Mediante el algoritmo de *Viterbi*, el primer modelo nos proporcionó la probabilidad de que un registro ECG, al que denominamos x , haya sido generado por un individuo perteneciente al grupo *mca*; es decir, nos proporciona la siguiente función de probabilidad:

$$\lambda_{mca}(x) = \hat{p}(x|mca) \quad (6.21)$$

El mismo proceso de decodificación utilizando el segundo modelo HMM se realizó para obtener la probabilidad de que el mismo ECG x provenga de algún individuo del grupo *control*. Esta probabilidad se denota con la siguiente expresión:

$$\lambda_{control}(x) = \hat{p}(x|control) \quad (6.22)$$

Finalmente, mediante el comando HVite de HTK, creamos un sistema de etiquetado que proporciona la clase más probable (*mca* o *control*) para cada registro ECG (x) durante la etapa de reconocimiento, denotado por la siguiente expresión:

$$\hat{c}(x) = \operatorname{argmax}(\lambda_{mca}(x), \lambda_{control}(x)) \quad (6.23)$$

Todos los ficheros de configuración necesarios para realizar esta experimentación, así como el código fuente de los *scripts* y programas desarrollados para ejecutar el flujo de trabajo están disponibles en https://github.com/sjimenezupv/htk_mca. Además, se incluyen los modelos HMM entrenados con todos los parámetros que dieron lugar a los resultados presentados

más adelante. Asimismo, también se proporcionan las herramientas creadas para extraer las métricas de precisión necesarias para evaluar el etiquetado en cada tupla de parámetros (nE, nG) de los modelos $\lambda_{\{mca, control\}}$ entrenados. Estos criterios de evaluación se describen a continuación.

6.3.10. Criterios de evaluación para los modelos ocultos de Markov

El objetivo de este estudio fue obtener modelos predictivos con un buen balance entre sensibilidad y especificidad. Para ello, estimamos una vez más su relación mediante la media geométrica $G = \sqrt{\text{Sensibilidad} \times \text{Especificidad}}$, definida en el intervalo $(0, 1)$.

Por otra parte, como se mencionó anteriormente, cada modelo HMM entrenado con HTK durante este estudio corresponde a una tupla de parámetros (nE, nG) . Para cada una de estas combinaciones, calculamos los valores de sensibilidad, especificidad y G utilizando el conjunto de *test*. Además, también se calculó la precisión con su intervalo de confianza (IC) al 5% de significancia estadística (α), evaluando con todas estas métricas el rendimiento de cada modelo de clasificación. Finalmente, fijamos un umbral de 0,7 en valores de G o de precisión indistintamente para considerar que un modelo HMM presentaba buenos resultados de clasificación.

6.4. Resultados

6.4.1. Análisis estadístico de todo el conjunto de datos

De las 87 variables disponibles, 30 resultaron estadísticamente significativas ($p < 0,01$). La Tabla 6.4 muestra sus valores, así como el valor p del test estadístico correspondiente, formando esta lista el conjunto total de datos D_t .

Tabla 6.4. Variables estadísticamente significativas (D_t).

D_t	MCA (n=41)	Control (n=46)	Valor p
<i>RAmp(I)</i>	0,402±0,423	0,716±0,397	8,08e-2
<i>RAmp(II)</i>	0,330±0,595	0,765±0,681	2,22e-3
<i>RAmp(aVR)</i>	-0,368±0,422	-0,762±0,347	7,50e-6
<i>RFRag(I)</i>	20 (48,8%)	10 (21,7%)	5,76e-4
<i>RFRag(aVR)</i>	20 (48,8%)	8 (17,4%)	1,75e-3
<i>RFRag(V5)</i>	14 (34,2%)	5 (10,9%)	8,71e-3
<i>Num_Rfrag</i>	6,659±2,921	4,804±2,156	1,05e-3
<i>QRSArea(II)</i>	0,900±9,879	5,762±6,844	8,55e-3
<i>QRSArea(aVR)</i>	-2,432±4,992	-5,377±4,036	3,15e-3
<i>midLevel(I)</i>	0,150±0,161	0,252±0,192	9,27e-2
<i>midLevel(II)</i>	0,094±0,258	0,314±0,327	8,52e-4
<i>midLevel(aVR)</i>	-0,122±0,169	-0,317±0,210	8,50e-6
<i>midLevel(aVF)</i>	-0,010±0,316	0,187±0,331	5,80e-3
<i>difLevel(I)</i>	0,319±0,298	0,520±0,363	6,47e-2
<i>difLevel(II)</i>	0,339±0,428	0,682±0,593	2,96e-3
<i>difLevel(aVR)</i>	0,296±0,293	0,638±0,417	3,44e-5
<i>difLevel(V3)</i>	0,437±0,458	0,828±0,579	8,59e-4
<i>TArea(I)</i>	2,585±5,550	9,349±6,219	8,00e-7
<i>TArea(II)</i>	4,513±7,282	10,508±8,060	4,94e-4
<i>TArea(V5)</i>	6,690±11,864	14,35±12,466	4,38e-3
<i>TArea(V6)</i>	3,724±9,143	11,767±10,181	2,21e-4
<i>TWAD</i>	0,249±0,466	0,591±0,209	2,10e-5
<i>TNeg(V6)</i>	9 (22,0%)	1 (2,2%)	3,89e-3
<i>Num_TNeg</i>	0,683±1,274	0,130±0,401	6,45e-3
<i>R2(VCG)</i>	0,653±0,210	0,783±0,227	6,88e-3
<i>D2(VCG)</i>	0,286±0,126	0,208±0,141	8,35e-3
<i>D2(VCG(QRS))</i>	-0,6275±0,6912	-1,0368±0,7314	9,00e-3
<i>RT(VCG(QRS))</i>	1,5939±0,3527	1,7859±0,2787	5,77e-3
<i>C1(PCA)</i>	0,036±0,021	0,022±0,015	1,00e-3
<i>C2(PCA)</i>	0,061±0,042	0,035±0,029	1,06e-3

Así mismo, la Figura 6.6 muestra los diagramas de cajas de las variables numéricas significativas en el conjunto de datos D_t extraídas a partir de valores de $midLevel$, $difLevel$, $Ramp$, $QRSArea$ y $TArea$ del ECG, distinguiendo entre el grupo de afectados por MCA-VI y el grupo control.

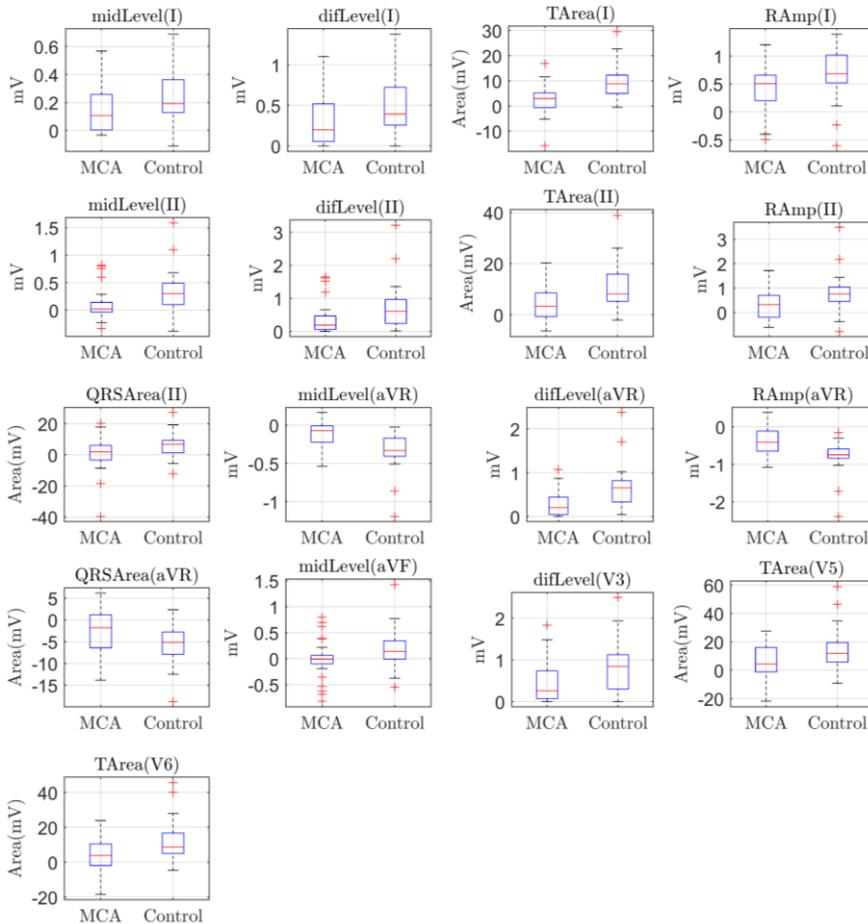


Figura 6.6. Diagramas de cajas de las variables numéricas significativas $midLevel$, $difLevel$, $Ramp$, $QRSArea$ y $TArea$ del conjunto de datos D_t .

En este análisis estadístico se observa que la MCA-VI se asocia significativamente a menor voltaje en las ondas R ($Ramp$) en las derivaciones I, II y aVR, así como a mayor área en el complejo QRS ($QRSArea$) en las derivaciones II y aVR. También se observa que la MCA-VI presenta menor área de las ondas T ($TArea$) en las derivaciones I, II, V5 y V6.

Por otro lado, hay un mayor número de ondas R fragmentadas ($RFRag$) en I, aVR, V5 y V6. La Figura 6.7 muestra diversos ejemplos de ondas R fragmentadas que aparecieron entre afectados por MCA en nuestro conjunto de datos. Además,

hay un mayor porcentaje de ondas T negativas (*TNeg*) en la derivación V6 entre los afectados por MCA-VI. La Figura 6.8 muestra dos ejemplos de ondas T negativas que detectamos en pacientes MCA en la base de datos empleada.

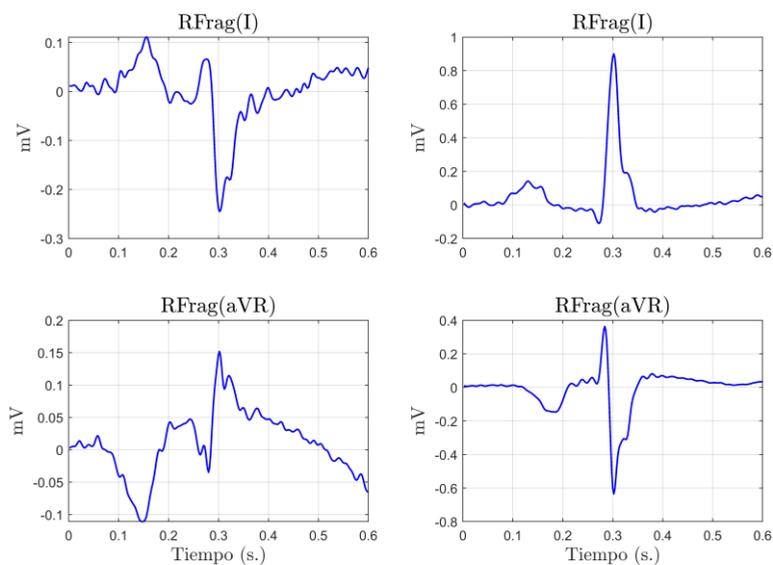


Figura 6.7. Detección de ondas R fragmentadas en las derivaciones I y aVR de cuatro pacientes afectados por MCA.

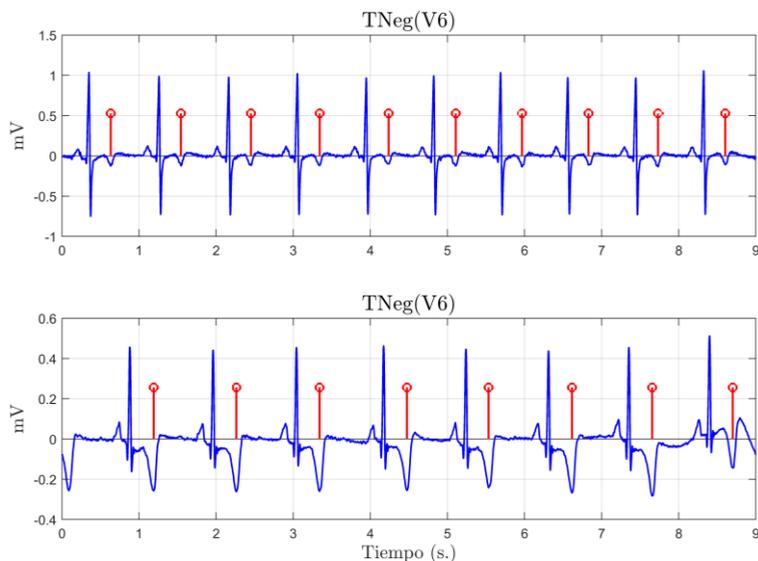


Figura 6.8. Detección de ondas T negativas en la derivación V6 de dos pacientes afectados por MCA.

Los resultados también mostraron diferencias significativas entre los grupos en el *midLevel* de las derivaciones I, II, aVR y aVF, y en el *difLevel* de las derivaciones I, II, aVR y V3. Las figuras Figura 6.9 y 6.10 ilustran los indicadores *midLevel* y *difLevel* tanto en un paciente MCA como en un control en la derivación I, donde aparecieron valores mayores en ambas variables en el grupo control. En dichas figuras, se aprecia como el *difLevel* se calcula como la diferencia entre el nivel superior y el inferior.

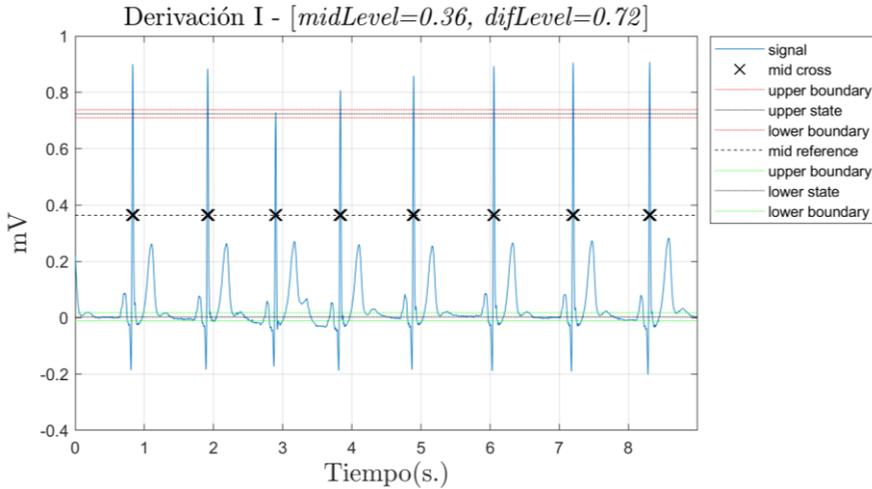


Figura 6.9. Ilustración de los indicadores *midLevel* y *difLevel* en la derivación I de un sujeto del grupo control.

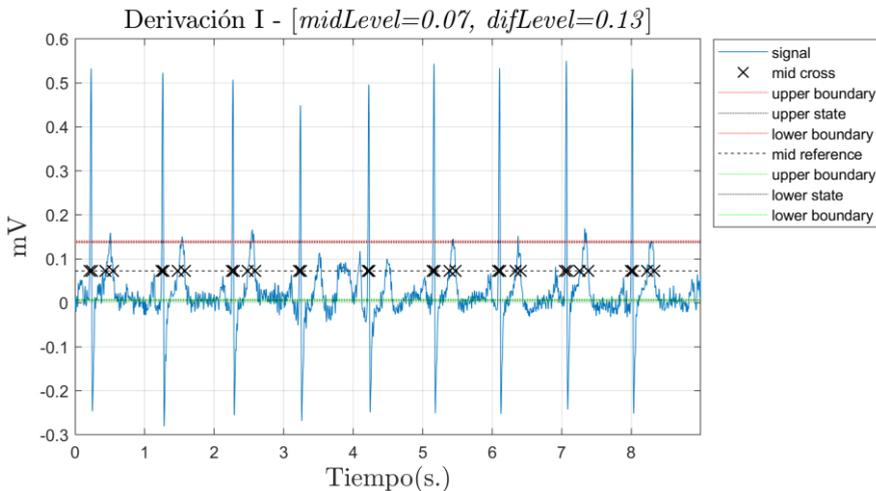


Figura 6.10. Ilustración de los indicadores *midLevel* y *difLevel* en la derivación I de un sujeto del grupo MCA.

Finalmente, la Figura 6.11 muestra los diagramas de cajas de las variables numéricas significativas en el conjunto de datos D_t extraídas a partir de valores del VCG, el PCA, la dispersión del área de la onda T (*TWAD*), así como el número total de ondas T negativas y ondas R fragmentadas. Todo ello distinguiendo entre el grupo de afectados por MCA-VI y el grupo control.

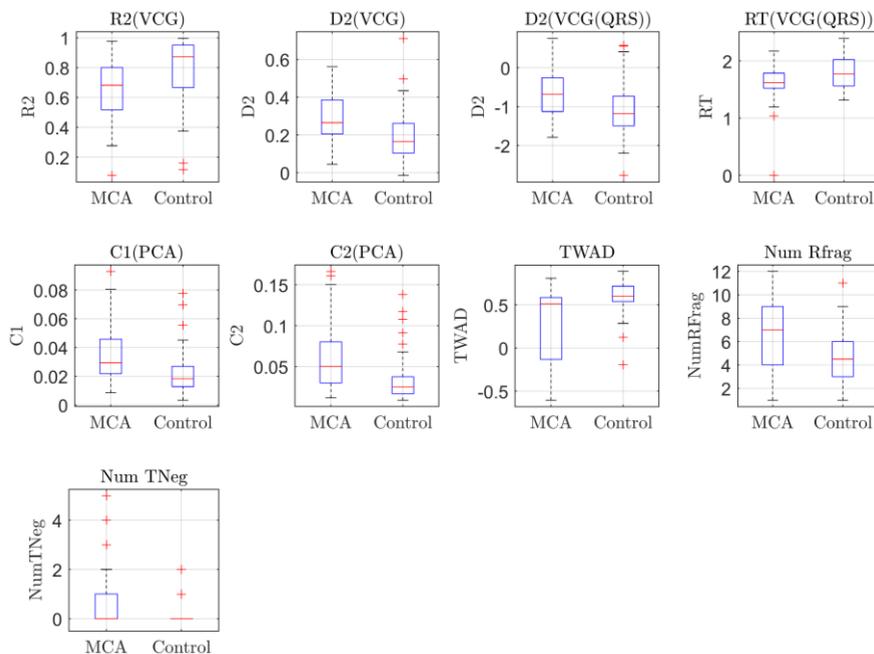


Figura 6.11. Diagramas de cajas de las variables numéricas significativas del conjunto de datos D_t extraídas de criterios del VCG, PCA, ondas T y número de ondas R fragmentadas.

El análisis estadístico mostró que tanto el número de derivaciones con la onda R fragmentada (*Num_RFrag*), como el número de ondas T negativas (*Num_TNeg*) aumentó significativamente en el grupo MCA. Respecto al área de las ondas T en distintas derivaciones (*TWAD*), se encontraron también diferencias significativas en su dispersión, concentrándose mucho más los valores entre aquellos del grupo control ($0,59 \pm 0,21$), mientras que los del grupo MCA por lo general eran inferiores a los anteriores en un rango más disperso ($0,25 \pm 0,47$).

Respecto las variables extraídas a partir del VCG, cuatro de ellas resultaron estadísticamente significativas. La primera fue el valor del estadístico R^2 de nivel de ajuste de cada VCG sobre un plano 2D ($R2(VCG)$), siendo este de $0,65 \pm 0,21$ y $0,78 \pm 0,23$ ($p < 0,01$) en el grupo de pacientes MCA y grupo control respectivamente, donde los mejores valores de ajuste se obtuvieron en el grupo control. La distancia normalizada $D2$ entre el VCG y su proyección sobre el plano de ajuste ($D2(VCG)$)

fue de $0,29 \pm 0,13$ y $0,21 \pm 0,14$ ($p < 0,01$) entre el grupo de pacientes y el de controles respectivamente. Esto indica que la distancia del VCG respecto al plano es mayor en el grupo de pacientes con MCA. Finalmente, los valores de los parámetros $D2(VCG(QRS))$ y $RT(VCG(QRS))$ también mostraron diferencias significativas, indicando patrones morfológicos distintos en el QRS y una trayectoria del VCG más compleja en el grupo de pacientes con MCA.

Las Figuras Figura 6.12, 6.13 y 6.14 corroboran mediante diferencias visuales los resultados mencionados respecto el VCG. En primer lugar, se observó que la mayoría de sujetos del grupo control de la base de datos poseen un VCG que no sufre demasiadas torsiones en el espacio 3D, estando muy próximos sus puntos al plano 2D de ajuste obtenido (ver Figura 6.12). Por otro lado, se observó que en gran parte de sujetos que padecen MCA, el VCG sufre más torsiones en el espacio 3D que en aquellos del grupo control. La Figura 6.13 muestra uno de estos casos, donde el plano que mejor se ajusta a los puntos del VCG queda alejado de éstos debido a la deformación sufrida.

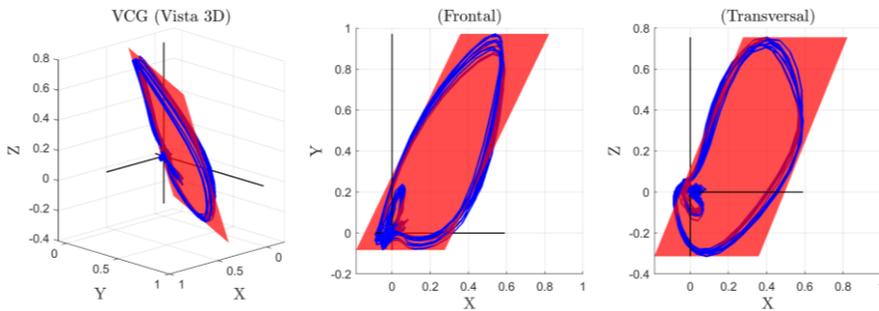


Figura 6.12. Vistas de un vectocardiograma de un sujeto del grupo control, junto con el plano que mejor se ajusta a sus puntos en el espacio 3D.

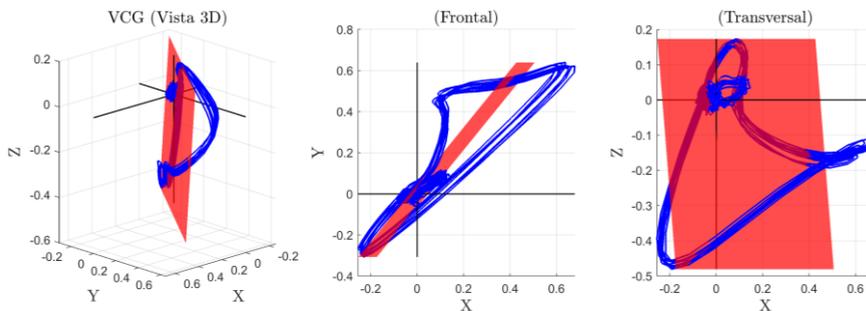


Figura 6.13. Vistas de un vectocardiograma de un sujeto afectado por MCA, junto con el plano que mejor se ajusta a sus puntos en el espacio 3D.

La Figura 6.14 muestra el lazo que forma el VCG del patrón QRS de un sujeto del grupo control (panel A), y también de un afectado por MCA (panel B). En ambos casos, además del eje de coordenadas 3D, también se muestra el plano que mejor se ajusta a sus puntos. Aquí podemos apreciar cómo la torsión en el QRS del afectado por MCA es evidente, mientras que el patrón del QRS en el sujeto control forma una elipse que puede ajustarse fácilmente al plano definido en el espacio 2D.

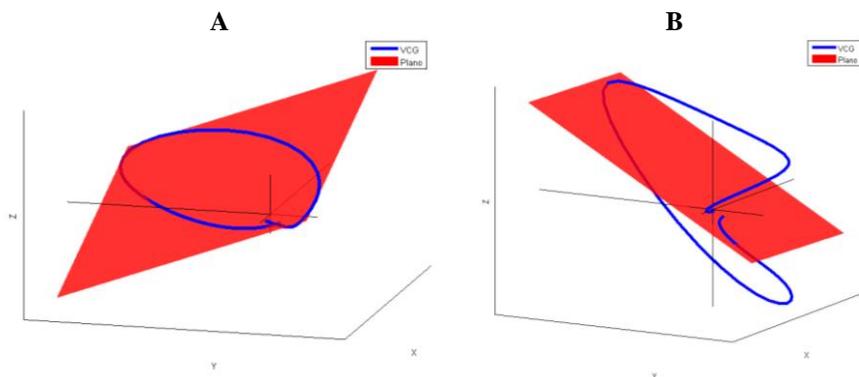


Figura 6.14. Vectocardiograma del patrón QRS de un sujeto del grupo control (A), y de un afectado por MCA (B).

Respecto a los indicadores extraídos a partir del PCA, obtuvimos dos indicadores estadísticamente significativos. El primero fue $C1(PCA)$, con valores de $0,036 \pm 0,021$ en pacientes con MCA y $0,022 \pm 0,015$ en los no afectados ($p < 0,01$). Esto indicó la existencia de más porcentaje de varianza en los componentes no dipolares entre los pacientes con MCA. Por otro lado, dentro del grupo control, los resultados sugieren que los componentes no dipolares contienen menos varianza, lo que es indicativo de una activación eléctrica más homogénea y organizada con respecto a los pacientes con afectación de MCA.

El segundo criterio significativo fue $C2(PCA)$, con valores de $0,061 \pm 0,042$ y $0,035 \pm 0,029$ ($p < 0,01$) en la misma división de grupos. Los valores de este indicador mostraron un mayor porcentaje de varianza en el tercer *eigenvalor* de los componentes dipolares entre los afectados por MCA. Este resultado es consistente con el ajuste del bucle del VCG a un plano, ya que, en el caso de los sujetos del grupo de control, se puede representar gran parte de la información sólo con dos componentes, mientras que en los pacientes con MCA hay una mayor presencia de la tercera componente que excede la restricción dimensional impuesta por un plano.

En resumen, ambos resultados en $C1(PCA)$ y $C2(PCA)$ presentan valores más elevados y heterogéneos en pacientes con MCA, lo que indica que la varianza se distribuye más en los últimos 6 *eigenvalores* en presencia de la enfermedad. Por otro lado, los dos primeros *eigenvalores* podrían explicar la mayor parte de la varianza

en el caso del grupo de control al estar los valores de ambos indicadores más concentrados en su media y más cercanos al cero.

6.4.2. Análisis estadístico estratificado por sexo

De las 87 variables disponibles, sólo 9 resultaron estadísticamente significativas ($p < 0,01$) para aquellas muestras correspondientes a personas del sexo femenino. Esto supuso el 10,3% de las variables disponibles; además, todas ellas también mostraron previamente diferencias significativas en el conjunto de datos D_t . La Tabla 6.5 muestra sus valores, así como el valor p del test estadístico correspondiente, formando esta lista el conjunto de datos D_{fem} . Así mismo, la Figura 6.15 muestra los diagramas de cajas de las variables numéricas significativas en el conjunto de datos D_{fem} extraídas a partir de valores del ECG y el PCA, distinguiendo entre afectados por MCA-VI y el grupo control.

En el análisis estadístico del conjunto de datos femenino se observa que la MCA-VI se asocia significativamente a menor voltaje ($RAmp$) y mayor fragmentación ($Rfrag$) en las ondas R de la derivación aVR. También se encontró que la MCA-VI presenta menor área de las ondas T ($TArea$) en las derivaciones I y V6, así como valores más heterogéneos en el parámetro $TWAD$, lo que indica mayor variabilidad en dicha onda T para este grupo.

Además, al igual que en el conjunto de datos D_t , hay un mayor porcentaje de ondas T negativas ($TNeg$) en la derivación V6 entre las mujeres afectadas por MCA-VI. Los resultados también mostraron diferencias significativas en los parámetros $midLevel$ y $difLevel$ en aVR, siendo esta derivación la que más aparece representada en el conjunto D_{fem} . Respecto a los indicadores extraídos a partir del PCA, el criterio $C2(PCA)$, resultó estadísticamente significativo con valores inferiores y más homogéneos en el grupo control.

Tabla 6.5. Variables significativas para muestras del sexo femenino (D_{fem}).

D_{fem}	MCA (n=18)	Control (n=28)	Valor p
$RAmp(aVR)$	-0,358±0,443	-0,689±0,201	1,22e-4
$Rfrag(aVR)$	9 (50,0%)	4 (14,3%)	8,65e-3
$midLevel(aVR)$	-0,131±0,182	-0,303±0,136	6,48e-4
$difLevel(aVR)$	0,330±0,305	0,609±0,273	2,35e-3
$TArea(I)$	1,397±5,909	7,509±4,561	2,80e-4
$TArea(V6)$	1,038±9,327	8,484±5,787	1,67e-3
$TNeg(V6)$	4 (22,2%)	0 (0,0%)	9,04e-4
$TWAD$	0,188±0,484	0,591±0,247	5,29e-4
$C2(PCA)$	0,073±0,049	0,041±0,032	1,00e-2

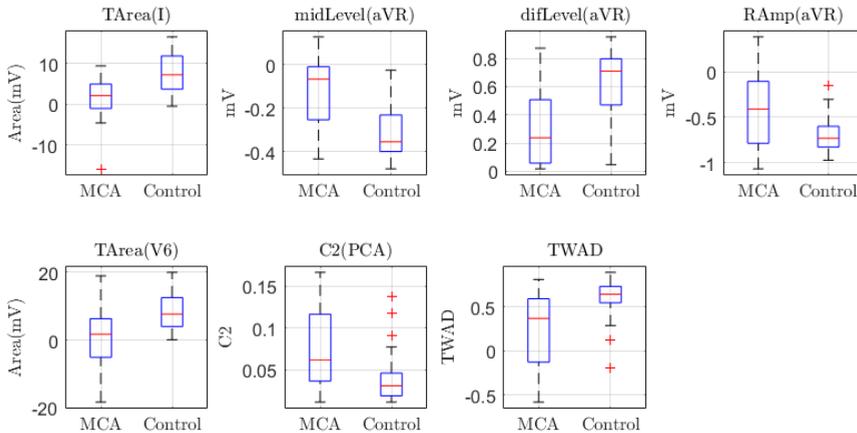


Figura 6.15. Diagramas de cajas de las variables numéricas significativas en el conjunto de datos D_{fem} extraídas de criterios del ECG y PCA.

Por otro lado, para las muestras correspondientes a personas del sexo masculino, de las 87 variables disponibles, 15 resultaron estadísticamente significativas ($p < 0,01$) lo que supuso un 17,2% de las variables disponibles. También en este caso, todos los parámetros mostraron previamente diferencias significativas en el conjunto de datos D_t . La Tabla 6.6 muestra sus valores, así como el valor p del test estadístico correspondiente, formando esta lista el conjunto de datos D_{mas} .

Tabla 6.6. Variables significativas para muestras del sexo masculino (D_{mas}).

D_{mas}	MCA (n=23)	Control (n=18)	Valor p
$RAmp(I)$	0,380±0,472	0,790±0,460	8,03e-3
$RAmp(aVR)$	-0,376±0,415	-0,875±0,483	1,00e-3
$QRSArea(aVR)$	-2,094±4,915	-6,661±4,828	5,00e-3
$midLevel(aVR)$	-0,115±0,161	-0,339±0,294	3,55e-3
$difLevel(aVR)$	0,269±0,288	0,683±0,581	4,88e-3
$difLevel(V3)$	0,375±0,391	0,969±0,648	7,96e-4
$TArea(I)$	3,514±5,192	12,211±7,429	7,82e-5
$TArea(II)$	4,890±7,561	15,366±8,702	1,91e-4
$TArea(V5)$	9,689±12,201	21,938±14,333	5,28e-3
$TArea(V6)$	5,826±8,618	16,874±13,259	2,55e-3
$R2(VCG)$	0,635±0,212	0,831±0,169	2,68e-3
$D2(VCG(QRS))$	-0,453±0,679	-1,087±0,670	4,83e-3
$RT(VCG(QRS))$	1,662±0,248	1,892±0,269	7,17e-3
$CI(PCA)$	0,035±0,021	0,019±0,013	4,92e-3
$C2(PCA)$	0,052±0,034	0,025±0,023	7,14e-3

Del mismo modo, la Figura 6.16 muestra los diagramas de cajas de las variables numéricas significativas en el conjunto de datos D_{mas} extraídas a partir de valores del ECG, VCG y PCA, distinguiendo entre afectados por MCA-VI y el grupo control.

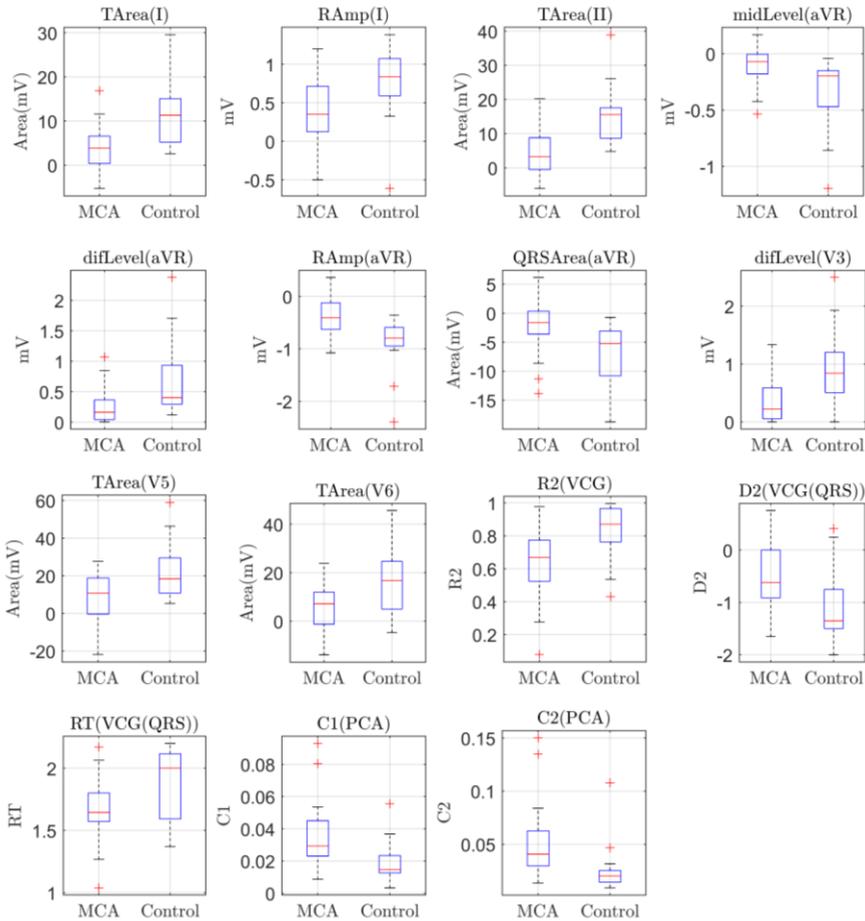


Figura 6.16. Diagramas de cajas de las variables numéricas significativas en el conjunto de datos D_{mas} extraídas de criterios del ECG, VCG y PCA.

En el análisis estadístico del conjunto de datos masculino se observa que la MCA-VI se asocia significativamente a menor voltaje en las ondas R ($RAmp$) en las derivaciones I y aVR, así como a mayor área en el complejo QRS ($QRSArea$) en la derivación aVR. Los resultados también mostraron diferencias significativas en los parámetros $midLevel$ y $difLevel$ en aVR, siendo una vez más la derivación más representada del conjunto de datos estratificado por sexo, en este caso D_{mas} . De

igual manera, *difLevel* en V3 también mostró diferencias significativas, donde los valores del grupo control en general fueron más elevados.

También se observa que la MCA-VI en hombres presenta menor área de las ondas T (*TArea*) en I, II, V5 y V6, siendo estas derivaciones exactamente las mismas que aparecen en el conjunto de datos D_t para este parámetro.

Respecto las variables extraídas a partir del VCG, tres de ellas resultaron estadísticamente significativas en la división de datos masculina: $R2(VCG)$, $D2(VCG(QRS))$ y $RT(VCG(QRS))$. Esto indicó que, tanto en el VCG como en el patrón VCG(QRS), aparecen patrones morfológicos y trayectorias más complejas en hombres afectados por la enfermedad.

Finalmente, los valores obtenidos en $CI(PCA)$ y $C2(PCA)$ también resultaron ser estadísticamente significativos, mostrando el mismo patrón que en los conjuntos de datos anteriores, con valores menores y distribuciones más homogéneas en el grupo control.

La Tabla 6.7 compara las variables que finalmente resultaron significativas en cada división de datos por sexos. Ahí se puede apreciar cómo los parámetros $RFRag(aVR)$, $TNeg(V6)$ y $TWAD$ muestra diferencias únicamente en el grupo de mujeres. Además, en el grupo femenino ninguna variable derivada del VCG mostró diferencias significativas, mientras que en el masculino tres de estas variables mostraron diferencias considerables en sus distribuciones.

Tabla 6.7. Comparación de las variables estadísticamente significativas (X) obtenidas en la estratificación de datos por sexo (femenino y masculino).

	D_{fem} (n=46)	D_{mas} (n=41)
$RAmp(I)$		X
$RAmp(aVR)$	X	X
$RFRag(aVR)$	X	
$midLevel(aVR)$	X	X
$difLevel(aVR)$	X	X
$difLevel(V3)$		X
$TArea(I)$	X	X
$TArea(II)$		X
$TArea(V5)$		X
$TArea(V6)$	X	X
$TNeg(V6)$	X	
$TWAD$	X	
$R2(VCG)$		X
$D2(VCG(QRS))$		X
$RT(VCG(QRS))$		X
$CI(PCA)$		X
$C2(PCA)$	X	X

Por otro lado, además de los tres parámetros basados en el VCG mencionados anteriormente, $RAmp(I)$, $TArea(II)$, $TArea(V5)$, y $CI(PCA)$ sólo mostraron diferencias significativas en la división de datos correspondiente al sexo masculino.

En resumen, de las 9 variables significativas halladas en D_{fem} , y las 15 en D_{mas} , sólo 6 coincidieron para ambos sexos, obteniendo más indicadores relevantes y diferencias en sus valores entre aquellos del sexo masculino.

6.4.3. Modelos de Regresión Logística y selección de características

La Tabla 6.8 detalla el rendimiento de clasificación en los distintos modelos basados en regresión logística, validados mediante $LOOCV$. Los mejores resultados de clasificación utilizando las muestras de ambos sexos (D_t) alcanzaron valores de G de 0,84 o 0,83 dependiendo del método de selección de características empleado. Aunque estos valores resultaron muy próximos, el nivel de especificidad fue mayor en el caso del modelo $D_t\lambda_{FS}$ con un valor de 0,85, siendo el valor de sensibilidad de 0,83 en ambos clasificadores $D_t\lambda_{FS}$ y $D_t\lambda_{BE}$.

Por otra parte, los resultados de los modelos que separaron las muestras por sexo muestran una clara mejora en la identificación de la MCA en pacientes masculinos, obteniendo en el mejor de los casos valores de G de 0,88, con una sensibilidad y especificidad de 0,87 y 0,89 respectivamente en el modelo $D_{mas}\lambda_{BE}$. En contraste, la identificación de la MCA en el sexo femenino alcanzó valores máximos de G de 0,81, con una sensibilidad y especificidad de 0,83 y 0,79 respectivamente, todo ello en el modelo $D_{fem}\lambda_{ZB}$.

Tabla 6.8. Resultados de clasificación ($LOOCV$) de los modelos de regresión logística.

Modelo	Sen.	Esp.	G	AUC	Precisión (IC 95%)
$D_t\lambda_{FS}$	0,83	0,85	0,84	0,81	0,84 (0,76-0,92)
$D_t\lambda_{BE}$	0,83	0,83	0,83	0,84	0,83 (0,75-0,91)
$D_{fem}\lambda_{FS}$	0,72	0,82	0,77	0,73	0,78 (0,66-0,90)
$D_{fem}\lambda_{BE}$	0,72	0,86	0,79	0,77	0,80 (0,68-0,92)
$D_{fem}\lambda_{ZB}$	0,83	0,79	0,81	0,80	0,80 (0,69-0,91)
$D_{mas}\lambda_{FS}$	0,74	0,94	0,84	0,80	0,83 (0,71-0,94)
$D_{mas}\lambda_{BE}$	0,87	0,89	0,88	0,85	0,88 (0,78-0,98)
D_t^*	0,85	0,83	0,84	-	0,84(0,76-0,92)

Finalmente, al utilizar la regla de clasificación D_t^* se obtuvo un valor de G de 0,84 respecto todo el conjunto de datos, idéntico que en el mejor de los casos al usar una única regresión logística en $D_t\lambda_{FS}$. No obstante, utilizando D_t^* los valores de sensibilidad y especificidad se invierten respecto $D_t\lambda_{FS}$, siendo éstos de 0,85 y 0,83 respectivamente. Al ser este valor de sensibilidad el más elevado teniendo en cuenta tanto hombres como mujeres, el modelo D_t^* sería el más indicado si se deseara identificar correctamente el mayor número de afectadas y afectados por MCA.

Una vez seleccionadas las características y evaluado el rendimiento de cada RL mediante *LOOCV*, se generó un clasificador final entrenado con todas las muestras del conjunto de datos correspondiente y características seleccionadas. Los valores de los mejores umbrales de clasificación (th) para cada una de estas regresiones logísticas finales, *i.e.* aquellos que maximizaban el valor de G , se muestran en la Tabla 6.9.

Tabla 6.9. Umbrales de clasificación de las regresiones logísticas.

Modelo	Umbral (th)
$D_t\lambda_{FS}$	0.50
$D_t\lambda_{BE}$	0.55
$D_{fem}\lambda_{FS}$	0.30
$D_{fem}\lambda_{BE}$	0.50
$D_{fem}\lambda_{ZB}$	0.29
$D_{mas}\lambda_{FS}$	0.55
$D_{mas}\lambda_{BE}$	0.43

Nótese que el rango de los umbrales es (0, 1), siendo 0,5 el valor deseable ya que indica un modelo totalmente balanceado en las distintas zonas de decisión. En nuestro caso, los umbrales obtenidos fueron de $0,5\pm 0,05$ en 7 de 9 de ocasiones, mostrando una separación entre grupos adecuada.

Las Figura 6.17 y 6.18 ilustran el uso de estos umbrales de clasificación en la salida de las regresiones logísticas finales $D_t\lambda_{FS}$ y $D_t\lambda_{BE}$, donde los valores de dichos umbrales fueron de 0,5 y 0,55 respectivamente. También puede apreciarse cómo las salidas de los regresiones logísticas pueden utilizarse como una puntuación de riesgo (probabilidad de MCA positiva), donde un valor próximo a 1 indica un riesgo elevado, mientras que valores cercanos a 0 indican un riesgo reducido. Nuestros modelos presentaron cierto grado de solapamiento entre ambos grupos en cada una de las zonas de decisión. No obstante, todas las muestras que superaron valores de 0,9 en las salidas de los distintos modelos resultaron ser siempre casos MCA, mientras que aquellas que no excedían valores de 0,05 resultaron ser siempre controles. Esto nos da una gran fiabilidad en dichas zonas, mientras que en las

restantes estaríamos siempre dentro de cierto grado de incertidumbre en la clasificación.

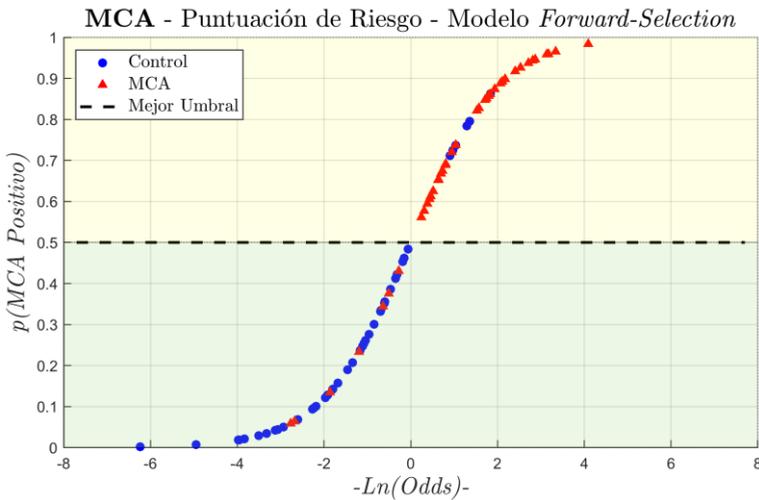


Figura 6.17. Valores de salida de la regresión logística $D_t\lambda_{FS}$, identificando pacientes, controles y el mejor umbral de decisión.

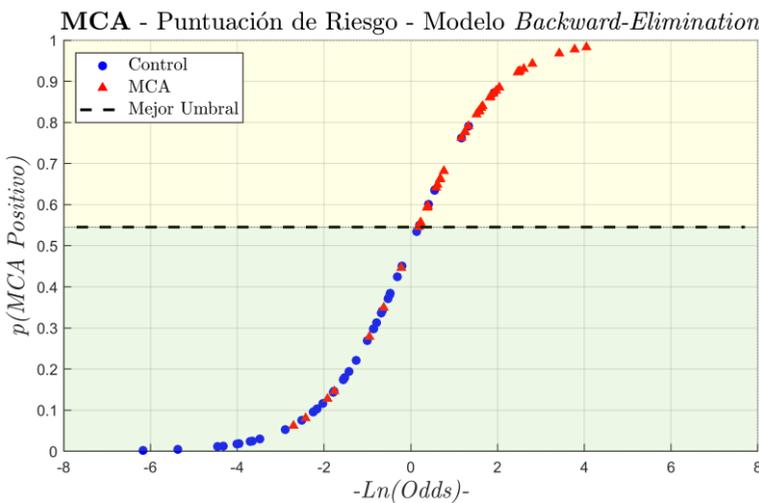


Figura 6.18. Valores de salida de la regresión logística $D_t\lambda_{BE}$, identificando pacientes, controles y el mejor umbral de decisión.

Respecto la selección de características, aparecieron combinaciones muy diversas de parámetros en cada modelo, sin apenas coincidencias entre ellos. En cuanto a los dos modelos que utilizaron las muestras de ambos sexos, tanto $D_t\lambda_{FS}$

como $D_t\lambda_{BE}$ seleccionaron los parámetros $TArea(I)$, $C1(PCA)$ y $difLevel(V3)$. Además, individualmente $D_t\lambda_{FS}$ seleccionó también los parámetros $D2(VCG)$, $R2(VCG)$, $midLevel(I)$, $RAmp(I)$ y $RFRag(I)$; mientras que $D_t\lambda_{BE}$ seleccionó también los parámetros $difLevel(I)$ y $RT(VCG(QRS))$. En total, $D_t\lambda_{FS}$ seleccionó 8 parámetros, a diferencia de $D_t\lambda_{BE}$, que seleccionó sólo 5, siendo el rendimiento de ambos clasificadores semejante.

Por otro lado, las variables seleccionadas en el conjunto de datos únicamente con muestras del sexo femenino fueron 2 para el modelo $D_{fem}\lambda_{FS}$ ($RAmp(aVR)$, $TNeg(V6)$), y 4 para los modelos $D_{fem}\lambda_{BE}$ ($difLevel(aVR)$, $RFRag(aVR)$, $C2(PCA)$, $TWAD$) y $D_{fem}\lambda_{ZB}$ ($TArea(I)$, $midLevel(aVR)$, $difLevel(aVR)$, $C2(PCA)$).

Finalmente, las regresiones logísticas que emplearon únicamente muestras del sexo masculino seleccionaron 2 parámetros ($C2(PCA)$, $TArea(V5)$) en el caso del modelo $D_{mas}\lambda_{FS}$, y 3 parámetros ($TArea(I)$, $D2(VCG(QRS))$, $C1(PCA)$) en el caso del modelo $D_{mas}\lambda_{BE}$. Además, se dio la circunstancia que el modelo $D_{mas}\lambda_{ZB}$ seleccionó exactamente los mismos parámetros que $D_{mas}\lambda_{BE}$, por lo que los resultados de clasificación del primero fueron idénticos, aunque el tiempo de cómputo para alcanzar la solución $D_{mas}\lambda_{ZB}$ fue muy superior al sufrir una explosión combinatoria al borde de ser computacionalmente inviable.

En resumen, 19 indicadores resultaron seleccionados en los distintos modelos RL entrenados, de los cuales 13 pertenecían al grupo de parámetros ECG: 9 relacionados con la amplitud o morfología de la onda R y 4 relacionados con el área de la onda T. De los 6 parámetros restantes, 4 estuvieron relacionados con el VCG y 2 con la caracterización del análisis PCA.

También cabe destacar que sólo 3 de entre los 19 indicadores resultaron seleccionados más de dos veces en distintos modelos de clasificación: $TArea(I)$, $C1(PCA)$ y $C2(PCA)$. Esto da una medida de la relevancia a la hora de detectar la MCA que provee el área de la onda T en la derivación I, así como los distintos criterios basados en la PCA definidos en este trabajo.

La Tabla 6.10 muestra una comparación de las variables seleccionadas para la clasificación por Regresión Logística en cada conjuntos de datos empleado (D_t , D_{fem} , D_{mas}) con los respectivos métodos de selección de característica (λ_{FS} , λ_{BE} , λ_{ZB}).

Tabla 6.10. Comparación de las variables seleccionadas (X) para la clasificación por Regresión Logística.

	$D_t\lambda_{FS}$	$D_t\lambda_{BE}$	$D_{fem}\lambda_{FS}$	$D_{fem}\lambda_{BE}$	$D_{fem}\lambda_{ZB}$	$D_{mas}\lambda_{FS}$	$D_{mas}\lambda_{BE}$ $D_{mas}\lambda_{ZB}$	Total
<i>TArea(I)</i>	X	X	-	-	X	-	X	4
<i>C1(PCA)</i>	X	X	-	-	-	-	X	3
<i>C2(PCA)</i>	-	-	-	X	X	X	-	3
<i>difLevel(V3)</i>	X	X	-	-	-	-	-	2
<i>RAmp(I)</i>	X	-	-	-	-	-	-	1
<i>RAmp(aVR)</i>	-	-	X	-	-	-	-	1
<i>midLevel(I)</i>	X	-	-	-	-	-	-	1
<i>RFRag(I)</i>	X	-	-	-	-	-	-	1
<i>RFRag(aVR)</i>	-	-	-	X	-	-	-	1
<i>midLevel(aVR)</i>	-	-	-	-	X	-	-	1
<i>difLevel(I)</i>	-	X	-	-	-	-	-	1
<i>difLevel(aVR)</i>	-	-	-	X	X	-	-	1
<i>TArea(V5)</i>	-	-	-	-	-	X	-	1
<i>TNeg(V6)</i>	-	-	X	-	-	-	-	1
<i>R2(VCG)</i>	X	-	-	-	-	-	-	1
<i>D2(VCG)</i>	X	-	-	-	-	-	-	1
<i>D2(VCG(QRS))</i>	-	-	-	-	-	-	X	1
<i>RT(VCG(QRS))</i>	-	X	-	-	-	-	-	1
<i>TWAD</i>	-	-	-	X	-	-	-	1
Total	8	5	2	4	4	2	3	

6.4.4. Modelos ocultos de Markov

La Tabla 6.11 muestra los mejores resultados de rendimiento de los modelos HMM obtenidos durante el proceso de validación en el conjunto de datos de *test*. Las columnas indican el número de estados utilizados (nE), el número de gaussianas (nG) de las distribuciones de probabilidad de emisión por estado b_j , así como las medidas de rendimiento en la clasificación: sensibilidad, especificidad, G , y precisión. Las filas se ordenan de mayor a menor valor de G . Podemos apreciar como 9 modelos superaron el criterio de obtener valores de G o precisión mayores que 0,7 para considerar que presentaban resultados aceptables de clasificación.

Dos modelos HMM diferentes alcanzaron el mayor balance entre sensibilidad y especificidad con un valor G de 0,83, y valores de sensibilidad y especificidad de 0,75 y 0,92 respectivamente en ambos casos. El primer modelo necesitó 30 estados y 4 gaussianas por estado, mientras que el segundo utilizó 40 estados y 4 gaussianas por estado.

El tercer mejor modelo HMM logró un valor G de 0,82 con el mayor nivel de especificidad posible (1,0) pero un bajo nivel de sensibilidad (0,67). La misma situación de sensibilidad y especificidad desequilibrada ocurre en los tres modelos siguientes, con valores de G entre 0,78 y 0,73.

Finalmente, tres modelos más alcanzaron valores de G de 0,72, donde uno de ellos obtuvo la sensibilidad más alta (0,83) durante toda la búsqueda de parámetros, utilizando 40 estados y 8 gaussianas por estado.

Tabla 6.11. Resultados de clasificación en el conjunto de datos de test de los HMM que superaron el umbral de 0,7 en valores de G o precisión.

nE	nG	Sen.	Esp.	G	Precisión (IC 95%)
30	4	0,75	0,92	0,83	0,84 (0,69-0,99)
40	4	0,75	0,92	0,83	0,84 (0,69-0,99)
45	4	0,67	1,00	0,82	0,84 (0,69-0,99)
25	4	0,67	0,92	0,78	0,80 (0,63-0,97)
35	4	0,67	0,92	0,78	0,80 (0,63-0,97)
50	4	0,58	0,92	0,73	0,76 (0,58-0,94)
40	8	0,83	0,62	0,72	0,72 (0,53-0,91)
55	4	0,75	0,69	0,72	0,72 (0,53-0,91)
60	8	0,75	0,69	0,72	0,72 (0,53-0,91)

Todos los modelos HMM mencionados anteriormente presentaron una precisión de 0,84 en los mejores casos, y 0,72 en los peores, mostrando siempre valores cercanos a su valor de G .

Las Figura 6.19 y 6.20 muestran los resultados de G y de precisión alcanzada durante toda la búsqueda de parámetros realizada mediante gráficos de pseudo-color, según el número de estados (nE) y gaussianas por estado (nG) utilizadas. En estas gráficas, el eje X o de abscisas corresponde al parámetro nE del HMM, mientras que el eje Y o de ordenadas corresponde al parámetro nG . El color indica los valores de G (Figura 6.19) o precisión (Figura 6.20) obtenidos para cada tupla (nE, nG) representando un modelo HMM evaluado.

En estas figuras se puede apreciar a simple vista que utilizar menos de 4 gaussianas por estado redujo significativamente el rendimiento de las predicciones, apareciendo zonas donde los valores de G tienden a cero, como en los modelos HMM que utilizan 2 gaussianas por estado. Esto indica que son necesarias mezclas de más de 2 gaussianas en las distribuciones de probabilidad de emisión b_j para conseguir un correcto aprendizaje. Por otro lado, utilizar más de 8 gaussianas provocó un sobre-ajuste de los modelos a los datos de entrenamiento, reduciendo su rendimiento en las predicciones del conjunto de *test*.

Los mejores resultados aparecen en el cuadrante comprendido entre 25 y 60 estados y de 4 a 8 gaussianas por estado, tal y como confirma la Tabla 1. Esta zona presenta un patrón homogéneo de modelos con buen rendimiento, donde gran parte de ellos superaron el umbral de 0,7 en G o en precisión, superando en algunos casos valores de 0,8.

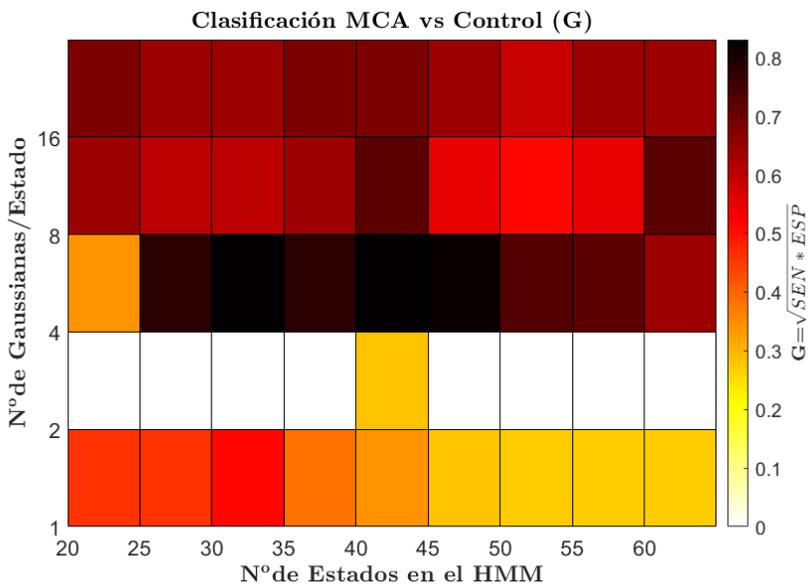


Figura 6.19. Valores de G obtenidos por los HMM en el conjunto de datos de *test*.

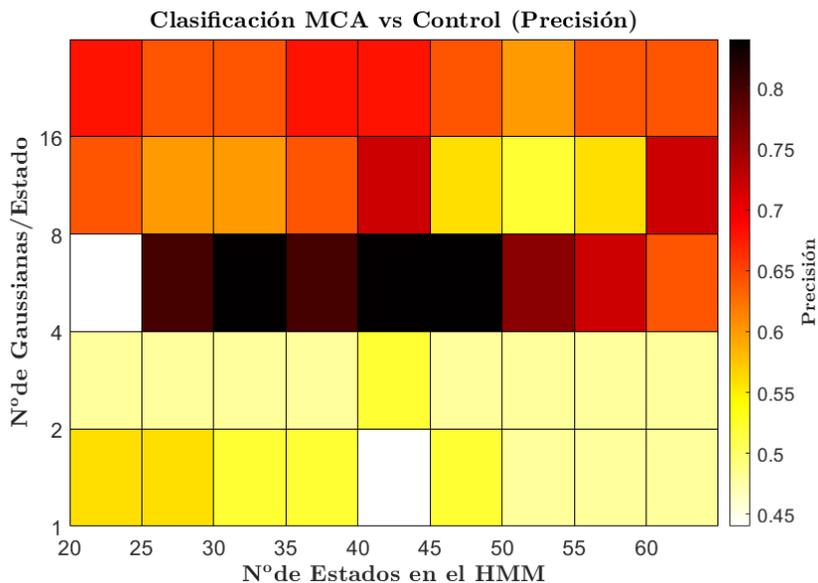


Figura 6.20. Valores de precisión obtenidos por los HMM en el conjunto de datos de test.

6.5. Discusión

En este estudio se han propuesto un conjunto de indicadores y modelos de clasificación con el fin de asistir al diagnóstico de la MCA en sus formas de afectación VI y biV.

Debe tenerse en cuenta que la base de datos de este trabajo evolucionó durante tres años y medio, incorporando nuevos casos y controles progresivamente, y realizando en ella distintos estudios referenciados en la Tabla 6.12. Los siguientes apartados, además de discutir los resultados obtenidos, compararán la significancia estadística obtenida con los trabajos previos.

Tabla 6.12. Evolución y estudios realizados con la base de datos MCA

Estudio	Pacientes MCA (n)	Controles (n)
[Jiménez 2015] VCG en MCA	34 (25 VI, 9 biV)	37
[Jiménez 2016a] PCA en MCA	32 (23 VI, 9 biV)	37
[Jiménez 2016b] Clasificación HMM	32 (23 VI, 9 biV)	37
[Jiménez 2017b] Clasificación RL	39 (25 VI, 14 biV)	43
Trabajo actual	41 (25 VI, 16 biV)	46

6.5.1. Criterios diagnósticos de MCA en el ECG

En primer lugar, se observa que la afectación de la MCA con afectación VI y biV tiene algunas repercusiones en el ECG no recogidas en los actuales criterios diagnóstico. Encontramos 30 indicadores significativos, siendo los más importantes la dispersión del área de la onda T reflejada en el parámetro *TWAD*, la amplitud de la onda R en la derivación aVR, la fragmentación de las ondas R, así como las áreas del QRS y la onda T en distintas derivaciones. Además, uno de los principales hallazgos de este estudio es que existen diferencias en el comportamiento de los marcadores al estratificar, encontrando sólo 9 indicadores significativos en el caso del sexo femenino, mientras que en el caso del sexo masculino se encontraron 15. Entre las mujeres destaca el área y el signo de la onda T en la derivación V6, además del parámetro *TWAD*, mientras que de estos mismos parámetros en hombres sólo mostró diferencias el área de la onda T en V6. En ambos sexos, los parámetros extraídos de la derivación aVR resultaron especialmente relevantes.

También se ha demostrado que existen diferencias significativas en algunos indicadores propuestos que evalúan el grado de torsión de la trayectoria del VCG, tanto en el ECG como en el QRS, siendo el grado de torsión mayor en el grupo de afectados por MCA, principalmente en hombres. Comparando los resultados de este trabajo respecto aquellos obtenidos en [Jiménez 2015], continuamos obteniendo significancia estadística en *R2* y *D2* a partir del VCG y del VCG(QRS). No obstante, se perdió dicha significancia en los parámetros *D1* y *S*, lo que mostró la importancia de aumentar el número de muestras para descartar criterios que inicialmente dieron buenos resultados en el trabajo anterior.

Asimismo, se hallaron diferencias significativas entre el grupo de pacientes y el de control en los criterios basados en los *eigenvalores* del PCA extraídos del ECG. Los resultados indicaron que el grado de información contenido en las derivaciones no dipolares (de la cuarta a la octava derivación) del PCA, reflejado en el criterio *C1(PCA)*, es mayor en el grupo de afectados por MCA. Además, el nivel de información contenida en la tercera dimensión respecto a las dos primeras, reflejado en el criterio *C2(PCA)*, también es mayor en el grupo de afectados por MCA. Todo esto indica una mayor varianza de los *eigenvalores* en el grupo de afectados por MCA, lo que sugiere una actividad eléctrica más descoordinada en presencia de la enfermedad.

Además, las diferencias en el indicador *C2(PCA)* también se mantuvieron al estratificar por sexo, tanto en hombres como en mujeres, mientras que el indicador *C1(PCA)* mantuvo sus diferencias en el subconjunto masculino.

Comparando los resultados de este trabajo respecto aquellos obtenidos en [Jiménez 2016a], al aumentar el número de muestras de la base de datos se mantuvo la tendencia en las distribuciones y la significancia estadística de los parámetros *C1(PCA)* y *C2(PCA)*. Esto sugiere que dichos marcadores presentan un

comportamiento robusto en la separación de los grupos de estudio, incluso al estratificar los datos entre los distintos sexos.

Todos los hallazgos en los indicadores propuestos son consistentes con los criterios diagnósticos actuales, donde se incluye la fragmentación del QRS o la presencia de onda epsilon entre otros. Dichos criterios sugieren un mayor nivel de heterogeneidad de la activación eléctrica en pacientes con MCA producido principalmente por las infiltraciones fibroadiposas provocadas por esta enfermedad [Marcus 2010, Quarta 2012, Pilichou 2016]. Esto también se ve reflejado en las alteraciones del VCG, o en los *eigenvalores* de la PCA, así como en los indicadores basados en la amplitud y morfología del ECG propuestos.

6.5.2. Modelos de Regresión Logística

Otro de los principales hallazgos de este estudio es que los modelos de clasificación basados en regresión logística utilizados son útiles en la detección de MCA en cribados familiares. Esta aproximación mostró buenos niveles de sensibilidad y especificidad durante la clasificación de registros ECG entre el grupo de pacientes afectados por MCA y el grupo control. En este sentido, los rendimientos máximos de clasificación alcanzados en la métrica G fueron de 0,84.

Por otro lado, al igual que en [Pujadas 2022] donde se identifica la FA con distintos modelos separando hombres y mujeres, la clasificación de MCA de este estudio estratificando por sexos también mostró mejores resultados en pacientes masculinos. Aquí, alcanzamos valores máximos de G de 0,81 y 0,88 en mujeres y hombres respectivamente. Esto sugiere que la identificación de la enfermedad es más problemática en el sexo femenino.

Respecto a la carga computacional, el proceso de validación *LOOCV* empleado requiere que el modelo sea reajustado y validado tantas veces como muestras existan en el conjunto de datos. Aunque esto supone una desventaja en muchos otros modelos de clasificación, nuestro ajuste de parámetros utilizando un máximo de 86 muestras en cada iteración no supuso una elevada carga computacional, teniendo en cuenta que el entrenamiento y validación tomó menos de 30 segundos en todos los modelos.

Por otro lado, los resultados de este trabajo mejoraron también respecto aquellos obtenidos en [Jiménez 2017b], pasando de valores de G de 0,7 a los actuales valores máximos de 0,84 utilizando en ambos casos toda la base de datos sin estratificar por sexo. Esta mejora se alcanzó aumentando el número de muestras y utilizando otras variables además de aquellas relacionadas con la amplitud del QRS, junto con mejoras en los métodos de selección de características.

Todo lo indicado muestra que la clasificación mediante regresión logística presenta un comportamiento robusto y adecuado en la detección de la MCA. Además, se demuestra que las combinaciones de nuevos parámetros basados en la amplitud y morfología del ECG, junto con aquellos extraídos del VCG y el PCA,

pueden resultar en la mejora del diagnóstico de MCA en sus formas de afectación VI y biV. De validarse en series de pacientes más grandes, sería posible desarrollar aplicaciones para escritorio o dispositivos móviles que se comunicaran con equipos de adquisición de ECG, con el objetivo de mejorar el diagnóstico de la enfermedad, así como los actuales criterios diagnósticos, con un bajo coste computacional.

6.5.3. Modelos ocultos de Markov

El último hallazgo de este estudio es que los modelos HMM también pueden ser útiles en la detección de MCA, principalmente en cribados familiares. Esta aproximación mostró buenos niveles de sensibilidad y especificidad. Durante la clasificación de registros ECG entre el grupo de pacientes afectados por MCA y el grupo control, los mejores rendimientos de clasificación alcanzados en la métrica G así como en la precisión fueron de 0,83 y 0,84 respectivamente.

No obstante, la carga computacional para el entrenamiento y validación de estos modelos es mucho más elevada que en el caso de la regresión logística. Es por ello que se utilizó aquí una estrategia de validación *hold-out* y no la *LOOCV* que utilizamos para las regresiones logísticas. Además, tampoco realizamos la estratificación por sexos en este caso ya que los conjuntos resultantes para entrenamiento y validación utilizando la estrategia *hold-out* pasaban a ser demasiado pequeños y los resultados no alcanzarían una relevancia estadística adecuada.

Finalmente, cabe destacar que el rendimiento de clasificación en este trabajo comparado respecto al obtenido en [Jiménez 2016b] mejoró sustancialmente, pasando de valores de G de 0,75 a 0,83. Para conseguir esta mejora, además de aumentar el número de muestras, la etapa de preprocesado de señal pasó de utilizar un filtro pasa banda entre 1 y 45 Hz en el trabajo inicial, a otro entre 0,7 y 70 Hz junto con un *notch* de 50 Hz. Esto sugiere que los HMM presentan un comportamiento robusto en la identificación de la enfermedad, pero muy sensible a la etapa de filtrado seleccionada. Incrementar todavía más el número de muestras aplicando el filtrado indicado hace que estos modelos sean candidatos prometedores a aumentar aún más su rendimiento de clasificación.

6.5.4. Limitaciones y trabajos futuros

Este estudio presenta algunas limitaciones a tener en cuenta. A continuación, presentamos una lista de ellas, así como algunas sugerencias de trabajos futuros al respecto.

En primer lugar, con el fin de obtener más significancia estadística, sería interesante ampliar la base de datos de registros ECG empleada mediante la inclusión de nuevos pacientes y controles, ya que actualmente contiene un número relativamente reducido de muestras. En segundo lugar, convendría validar los distintos indicadores y modelos de clasificación propuestos en otras series de

pacientes, preferentemente en un estudio multicéntrico. No obstante, se debe tener en cuenta que las muestras del estudio son difíciles de conseguir debido a la baja incidencia de la enfermedad y al requisito de utilizar como controles a familiares directos de los afectados por MCA en sus formas VI y biV.

Es también necesario desarrollar criterios que identifiquen mejor la enfermedad entre mujeres, ya que los descritos en este trabajo no permiten a las regresiones logísticas alcanzar un rendimiento tan elevado como el logrado en el sexo masculino.

Respecto a la clasificación de registros ECG mediante modelos HMM, utilizamos 42 series temporales diferentes extraídas del ECG. No obstante, sería interesante explorar una reducción de la dimensionalidad de los datos con el objetivo de reducir el coste computacional y mejorar el rendimiento de clasificación alcanzado. También pudiera ser recomendable utilizar una combinación de modelos HMM para mejorar los resultados obtenidos, intentando de este modo aumentar aún más la sensibilidad en la detección de MCA mediante el ECG.

Finalmente, a pesar que ha sido posible cuantificar el rendimiento de distintos criterios diagnósticos y modelos de clasificación para la identificación de la MCA, los enfoques propuestos no son capaces por sí solos de separar con una precisión exacta los sujetos que pertenecen al grupo de afectados del de control. Aunque esto puede ser debido en gran medida a que existen sujetos en el grupo control con otras afecciones cardiacas que pueden alterar el ECG, se necesita todavía una investigación más exhaustiva de otros parámetros y técnicas que permitan mejorar los resultados actuales.

6.5.5. Conclusiones

En este estudio propusimos una serie de indicadores extraídos de registros cortos de ECG que presentan diferencias significativas entre afectados por MCA (VI y biV) y el grupo control, todos ellos participantes en cribados intra-familiares. Los criterios diagnósticos que mostraron diferencias significativas sugieren una mayor heterogeneidad de la activación eléctrica en pacientes con MCA.

Además, demostramos que es posible identificar con un nivel aceptable de sensibilidad y especificidad la MCA utilizando estos mismos registros ECG mediante distintas aproximaciones basadas en RL y HMM, con valores máximos de G de 0,84 y 0,83 respectivamente en nuestro caso.

Finalmente, demostramos que existen diferencias tanto en los indicadores como en el rendimiento de la clasificación entre los distintos sexos, observando mejoras en la identificación de la enfermedad en pacientes masculinos, donde alcanzamos valores máximos de G de 0,88 utilizando RL.

En su conjunto, todos estos resultados podrían mejorar el diagnóstico de la MCA en sus formas de afectación VI y biV, especialmente dentro del contexto del cribado familiar, ayudando a identificar y caracterizar mejor la enfermedad de manera no invasiva y aumentando la sensibilidad en la detección mediante ECG.

Capítulo 7

Conclusiones

En este capítulo, se presentan las conclusiones globales de los tres estudios realizados en el marco de esta tesis, en relación con los objetivos principales y secundarios enunciados en el Capítulo 1. Finalmente, el capítulo concluye con pautas para trabajos futuros, destacando aspectos de esta tesis que pueden explorarse más a fondo.

7.1. Conclusiones

7.1.1. Respecto los objetivos principales

Objetivo Principal 1. *Desarrollar y validar metodologías robustas y de bajo coste computacional, capaces de identificar una amplia variedad de afecciones cardiacas a partir de distintas combinaciones de derivaciones ECG, haciendo énfasis en los escenarios de derivación única. Este objetivo quiere abordar la necesidad clínica de poder identificar automáticamente múltiples afecciones cardiacas usando nuevas tecnologías portátiles que trabajan con información electrocardiográfica limitada. También pretendemos evaluar los rendimientos de clasificación con segmentos cortos de ECG, ya que son los más utilizados en dispositivos móviles e inteligentes.*

En el Capítulo 4 se ha presentado un primer estudio donde se evaluaba una metodología de clasificación de registros ECG, de corta duración y una única derivación, en cuatro clases de ritmos cardiacos distintos: ritmo sinusal normal, fibrilación auricular, otros ritmos anómalos, y registros ruidosos. Para ello, combinamos la extracción de características del ECG basadas en la variabilidad del intervalo RR, la selección de conjuntos óptimos de características, y el entrenamiento y validación de diferentes modelos de clasificación. Los mejores resultados mostraron un rendimiento de clasificación en la métrica F_1 de 0,75, con

un coste computacional de 15,1 ms por muestra. Se observó un nivel óptimo de separación entre RSN y FA, pero poca sensibilidad en la identificación de otros ritmos anómalos distintos a la FA, así como de los registros ruidosos, lo que hizo bajar el rendimiento global de clasificación del sistema. En resumen, los modelos desarrollados en este primer estudio presentaron un bajo consumo computacional durante la clasificación, y un buen rendimiento en la identificación de FA. Por lo tanto, podrían ser candidatos adecuados para ser incluidos en sistemas que utilicen dispositivos portátiles para la monitorización cardíaca, o de gestión de pacientes, con el objetivo de ayudar en la detección temprana de arritmias auriculares en cribados iniciales.

En el Capítulo 5, un segundo estudio evaluó una metodología para la detección de múltiples afecciones cardíacas a través de registros cortos de ECG, aplicando la extracción y selección de características, junto a un enfoque de clasificación *One-vs-Rest* que combinó tres clasificadores binarios distintos. Los resultados de clasificación utilizando una o dos derivaciones ECG no estuvieron muy alejados de aquellos que usaron 12 derivaciones, mostrando costes computacionales más bajos y siendo más adecuados para dispositivos portátiles. Además, para algunas enfermedades cardíacas en concreto, el uso de una o dos derivaciones ECG mostró rendimientos de clasificación iguales o superiores que otras configuraciones con más derivaciones. Esto sugiere que, dependiendo de la enfermedad a detectar, el uso de conjuntos reducidos de derivaciones es suficiente para un buen rendimiento de clasificación.

Objetivo Principal 2. *Definir nuevos biomarcadores y clasificadores basados en el ECG para el análisis diagnóstico de la MCA que contribuyan a detectar nuevos casos durante cribados intra-familiares, ayudando así a reducir el número de muertes súbitas por esta enfermedad.*

En el Capítulo 6, en un tercer estudio, propusimos una serie de indicadores extraídos de registros cortos de ECG que mostraron diferencias significativas entre afectados por MCA (VI y biV) y el grupo control, todos ellos participantes en cribados intra-familiares. Estos criterios diagnósticos sugieren una mayor heterogeneidad de la activación eléctrica en pacientes con MCA. Además, demostramos que es posible identificar con un nivel aceptable de sensibilidad y especificidad la MCA utilizando registros ECG mediante distintas aproximaciones basadas en regresión logística y modelos ocultos de Markov, alcanzando rendimientos de clasificación elevados con valores en la métrica G de 0,84 y 0,83 respectivamente. Finalmente, demostramos que existen diferencias tanto en los indicadores como en el rendimiento de clasificación entre los distintos sexos, observando mejoras en la identificación de la enfermedad en hombres, donde alcanzamos valores de G de 0,88.

En su conjunto, todos estos resultados podrían mejorar el diagnóstico de la MCA en sus formas de afectación VI y biV, especialmente dentro del contexto del cribado familiar, ayudando a identificar y caracterizar mejor la enfermedad de manera no invasiva y aumentando la sensibilidad en la detección mediante ECG. No obstante, estos indicadores y modelos de clasificación no deben ser utilizados individualmente, siendo necesario combinarlos con otros criterios clínicos para asegurar los resultados del diagnóstico.

7.1.2. Respecto los objetivos secundarios

Objetivo Secundario 1. *Desarrollar algoritmos para la extracción y selección automática de características temporales y espectrales derivadas de la actividad ventricular de las señales ECG disponibles.*

En el estudio del Capítulo 4, extrajimos automáticamente 81 características de la señal ECG basadas en la variabilidad del intervalo RR, requiriendo un bajo coste computacional medio de 13,7 ms por muestra. Uno de los principales hallazgos fue que las características pNN50, pNN75 y aquellas basadas en distancias a los vectores $pNNx$ mostraron un nivel de separación óptimo entre registros de las clases RSN y FA. Sin embargo, este grupo de características no mostró diferencias entre el RSN y la clase correspondiente a otras patologías; lo mismo sucedió con la FA y los registros ruidosos, solapándose su rango de valores mutuamente.

En este mismo estudio, desarrollamos un método de selección de características basado en *Forward-Selection* que consiguió reducir en un 84% el número de variables manteniendo el rendimiento de clasificación. En general, las características seleccionadas también presentaron un buen nivel de separación entre las clases RSN y FA, pero mostraron solapamientos con las demás clases. La excepción fueron las siguientes cuatro características que mostraron diferencias entre los valores correspondientes a registros ruidosos de los de FA: (1) desviación media de RRd1 en el conjunto de ventanas deslizantes; (2) desviación estándar entre todas las distancias de los puntos del gráfico de Poincaré; (3) distancia mínima entre los puntos del gráfico de Lorenz; (4) ratio de patrones QRS frente al número de ondas R detectadas.

En el estudio del Capítulo 5, extrajimos automáticamente 81 características de cada una de las 12 derivaciones ECG disponibles, además de la edad y el sexo de los metadatos de los registros, obteniendo un total de 974 valores de características por cada muestra. Estas características, se derivaron de la actividad ventricular, basándose principalmente en la variabilidad del ritmo cardíaco, patrones morfológicos de las ondas QRS y T, además de la parte baja del dominio espectral. Aquí, el coste computacional medio necesario para realizar el procesado de señal y

la extracción de características para cada registros ECG de 12 derivaciones fue de 9,52 ms por muestra.

El método de selección de características empleado, basado en filtrado estadístico supervisado y no supervisado, redujo en un 40% el número de variables, repartiéndolas de manera equilibrada entre las distintas derivaciones ECG en cada modelo de clasificación binario. Demostramos que nuestro método aliviaba la carga computacional que supondría entrenar y validar los modelos de clasificación con todas las características disponibles, manteniendo además el rendimiento de clasificación.

Finalmente, hallamos dos categorías de características que destacaron sobre el resto durante el proceso de selección. La primera fue la relacionada con las estadísticas temporales del intervalo QT, con una media de selección del 86% de las veces en el cómputo global. Además, tuvo una selección del 100% para los clasificadores binarios correspondientes a las siguientes patologías: desviación del eje a la derecha, mala progresión de la onda R, y *flutter* auricular. La segunda categoría de características que destacó en este estudio fue la relacionada con las estadísticas de los voltajes de las ondas R y T, con una media de selección del 75%.

En resumen, las características mencionadas deben ser tenidas en cuenta como posibles indicadores a incluir en dispositivos móviles para el diagnóstico y cribado, ya que son útiles en la evaluación de distintas cardiopatías, mostrando además buenos rendimientos computacionales.

Objetivo Secundario 2. *Entrenar y validar el rendimiento de diversos modelos de clasificación que etiqueten registros cortos de ECG de una única derivación entre los siguientes cuatro grupos: ritmo normal, Fibrilación Auricular, otros ritmos y ruido.*

En el estudio del Capítulo 4, utilizando 8.528 registros de ECG de una única derivación, entrenamos y validamos tres tipos de modelos de clasificación: FFNN, SVM y NB. Hallamos que los modelos con mejor rendimiento de clasificación y computacional fueron las FFNN (F_1 : 0,75; 1,4 ms/muestra), seguidas por los modelos SVM (F_1 : 0,73; 4,5 ms/muestra), mientras que NB consiguió rendimientos inferiores (F_1 : 0,67; 18,1 ms/muestra). Además, hallamos mayor dificultad en la identificación de las clases referentes a otros ritmos anómalos distintos a la FA, y de registros ruidosos, sugiriendo centrar los esfuerzos de futuros trabajos en mejorar su clasificación. En conclusión, las FFNN fueron los modelos que mejores rendimientos computacionales y de clasificación presentaron en este estudio, lo que las hacen candidatas idóneas para ser implementadas en dispositivos móviles de bajo consumo energético.

Objetivo Secundario 3. *Entrenar y validar el rendimiento de diversos modelos de clasificación que utilicen conjuntos de distintas derivaciones ECG, para etiquetar nuevas muestras entre 26 ritmos cardiacos diferentes, patológicos principalmente.*

En el estudio del Capítulo 5, utilizamos un enfoque de clasificación *one-vs-rest*, combinando técnicas de aprendizaje automático supervisadas y no supervisadas, y clasificadores binarios de tipo FFNN, NB o híbrido. Esta aproximación mostró buenos rendimientos de clasificación en una base de datos compuesta por 131.149 muestras de ECG de 12 derivaciones. En concreto, los valores medios en la métrica G durante la validación cruzada en el conjunto de entrenamiento público fueron de 0,80, 0,78, 0,79, 0,79, 0,78 y 0,74 usando combinaciones de 12, 6, 4, 3, 2 y 1 derivación ECG.

Como era de esperar, el mejor rendimiento se alcanzó utilizando 12 derivaciones, lo que sugiere que este estándar no debe reemplazarse por sistemas de derivaciones reducidos en la práctica clínica. Sin embargo, la disminución del rendimiento con menos derivaciones fue pequeña. Estos resultados muestran un gran potencial para conjuntos reducidos de derivaciones ECG fuera del entorno clínico, como por ejemplo, aquellos destinados a cribados y seguimientos masivos para la detección temprana de anomalías.

Por otro lado, cabe destacar que las clasificaciones que utilizaron 3 y 4 derivaciones superaron en rendimiento a aquellas que usaron 6. Esta circunstancia podría explicarse por el hecho de que, en el subconjunto de 6, donde se utilizan sólo tres electrodos, únicamente dos derivaciones son independientes. Asimismo, los subconjuntos de 3 y 4 derivaciones contienen la derivación precordial V2, la cual es independiente de las derivaciones de las extremidades y, por lo tanto, agrega información clínicamente valiosa. Estos resultados sugieren que es más importante incluir derivaciones con información mutuamente complementaria, que aumentar su número.

Las diferencias entre los resultados obtenidos utilizando únicamente la derivación I y aquellos obtenidos con el conjunto completo de 12 derivaciones fueron relativamente pequeñas, con valores medios de G de 0,74 y 0,80 respectivamente. Esto resulta de gran interés, ya que utilizar una única derivación posibilita el registro de ECG y el diagnóstico automatizado en diversas situaciones diarias donde el estándar de 12 derivaciones no es viable o posible, a pesar del hecho de que el rendimiento es menor con una derivación. Esto es así, por ejemplo, usando un dispositivo portátil o un reloj inteligente.

Además, cabe destacar que, en seis afecciones cardiacas (AS, FA, PRPr, BAVPG, BradiS, TacS), el rendimiento de la clasificación utilizando únicamente la derivación I igualó o mejoró la clasificación con 12 derivaciones. La mayoría de estos rendimientos incluso mejoraron al utilizar el modelo de clasificación de 2 derivaciones ECG.

Sin embargo, este estudio también mostró que algunas afecciones cardiacas necesitan información de derivaciones precordiales para mejorar su detección, como BR, DEjI, BFAnI, VQRSB, MPrR y DEjD, donde la configuración de 12 derivaciones superó claramente a las demás.

Desde un punto de vista computacional, el tiempo medio de clasificación con una única derivación ECG redujo a la mitad el tiempo medio de clasificación con 12 derivaciones: 56 ms/muestra utilizando una derivación ECG; 143 ms/muestra utilizando 12 derivaciones. Por lo tanto, los dispositivos portátiles e inteligentes saldrían beneficiados de unos menores costes computacionales y de consumo de batería al utilizar sólo una derivación.

Objetivo Secundario 4. *Definir y evaluar nuevos indicadores para el diagnóstico de MCA mediante registros ECG, analizando las diferencias existentes al estratificar por sexo.*

El primer hallazgo del estudio del Capítulo 6 es que la MCA con afectación VI y biV tiene algunas repercusiones en el ECG no recogidas en los actuales criterios diagnósticos. Encontramos 30 indicadores significativos, siendo los más importantes la dispersión del área de la onda T reflejada en el parámetro *TWAD*, la amplitud de la onda R en la derivación aVR, la fragmentación de las ondas R, así como las áreas del QRS y la onda T en distintas derivaciones. Además, se observaron diferencias en el comportamiento de los marcadores al estratificar por sexo, encontrando 15 indicadores significativos para el sexo masculino, y sólo 9 para el sexo femenino. Entre las mujeres, destaca el área y el signo de la onda T en la derivación V6, además del parámetro *TWAD*, mientras que, de estos parámetros en hombres, sólo mostró diferencias el área de la onda T en V6. En ambos sexos, los parámetros extraídos de la derivación aVR resultaron especialmente relevantes.

También encontramos diferencias significativas en algunos indicadores que evalúan el grado de torsión de la trayectoria del VCG, siendo el grado de torsión mayor en el grupo de afectados por MCA, principalmente en hombres.

Asimismo, se hallaron diferencias significativas en los criterios basados en los *eigenvalores* del PCA extraídos del ECG, donde apareció una mayor varianza de los valores en el grupo de afectados por MCA, sugiriendo una actividad eléctrica más descoordinada en presencia de la enfermedad.

Todos los hallazgos en los indicadores propuestos son consistentes con los criterios diagnósticos actuales, donde se incluye la fragmentación del QRS o la presencia de onda epsilon entre otros. Dichos criterios también sugieren un mayor nivel de heterogeneidad de la activación eléctrica en pacientes con MCA producido principalmente por las infiltraciones fibroadiposas provocadas por la enfermedad [Marcus 2010, Quarta 2012, Pilichou 2016]. Esto también se ve reflejado en las

alteraciones del VCG, o en los *eigenvalores* de la PCA, así como en los indicadores basados en la amplitud y morfología del ECG propuestos.

En resumen, las combinaciones de nuevos parámetros basados en la amplitud y morfología del ECG, junto con aquellos extraídos del VCG y el PCA, pueden resultar en la mejora del diagnóstico de MCA en sus formas de afectación VI y biV.

Objetivo Secundario 5. *Evaluar distintos modelos de clasificación, que identifiquen a afectados y no afectados por MCA utilizando registros cortos de ECG de 12 derivaciones, contemplando también la estratificación por sexo.*

Uno de los principales hallazgos del estudio del Capítulo 6 es que los modelos de clasificación basados en regresión logística son útiles en la detección de MCA en cribados familiares. Esta aproximación mostró buenos niveles de sensibilidad y especificidad en la detección de MCA en registros ECG, alcanzando rendimientos de clasificación en la métrica G de 0,84. Por otro lado, la identificación de MCA estratificando por sexos mostró mejores resultados en pacientes masculinos, alcanzando valores máximos de G de 0,81 y 0,88 en mujeres y hombres respectivamente. Esto sugiere que la detección de la enfermedad es más problemática en el sexo femenino. Respecto a la carga computacional, el proceso de entrenamiento y validación de las regresiones logísticas, utilizando 86 muestras, no supuso un coste elevado, necesitando menos de 30 segundos para los dos procesos. Todo lo indicado sugiere que la clasificación mediante regresión logística presenta un comportamiento robusto y adecuado en la detección de la MCA. De validarse en series de pacientes más grandes, sería posible desarrollar aplicaciones para escritorio o dispositivos móviles que se comunicaran con equipos de adquisición de ECG, con el objetivo de mejorar el diagnóstico de la enfermedad, así como los actuales criterios diagnósticos, con un bajo coste computacional.

El último hallazgo del estudio del Capítulo 6 es que los modelos ocultos de Markov también pueden ser útiles en la detección de MCA, principalmente en cribados familiares. Esta aproximación mostró buenos niveles de sensibilidad y especificidad, con rendimientos de clasificación en la métrica G de 0,83. También encontramos que los HMM presentan un comportamiento muy sensible a la etapa de filtrado de señal, variando su rendimiento de clasificación sustancialmente según los parámetros de filtrado seleccionados. No obstante, la carga computacional para el entrenamiento y validación de estos modelos es mucho más elevada que en el caso de la regresión logística. Incrementar el número de muestras para el entrenamiento y aplicar el filtrado adecuado son las direcciones adecuadas para aumentar más el rendimiento de clasificación de estos modelos.

7.2. Directrices para trabajos futuros

Los trabajos futuros en la detección de FA en señales ECG de una única derivación deben centrarse principalmente en el aumento de la sensibilidad en la identificación de ritmos anómalos distintos a la propia FA, con el fin de alcanzar un rendimiento global del sistema óptimo. Para ello se proponen las siguientes directrices: 1) Añadir características del ECG no relacionadas únicamente con la variabilidad del intervalo RR (amplitud y morfología del P-QRS-T, detección de latidos ectópicos, ausencias y/o inversiones de las ondas T y P, tiempo-frecuencia de la señal). 2) Abordar otras técnicas de selección de características, como los algoritmos genéticos, o el enfriamiento simulado. 3) Ampliación del conjunto de entrenamiento para conseguir una mayor generalización en la clasificación. 4) Explorar otros tipos de modelos de clasificación, principalmente basados en redes convolucionales, o modelos ocultos de Markov.

Por otro lado, los trabajos futuros referentes a la detección de múltiples afecciones cardíacas en diversos conjuntos de derivaciones deben centrarse, principalmente, en la inclusión de características y/o reglas clínicas más específicas para cada una de las afecciones cardíacas a detectar, especialmente en aquellas donde nuestros rendimientos de clasificación fueron menores. También debe abordarse la inclusión de otros tipos de clasificadores, como los árboles de decisión, sistemas de reglas, o redes convolucionales.

Respecto los trabajos futuros sobre indicadores y clasificadores para detectar MCA en el ECG, deben contemplar en primer lugar la ampliación de las bases de datos empleadas con el mayor número de registros posibles, tanto de nuevos pacientes como de controles adecuados, preferiblemente familiares de afectados. En segundo lugar, los distintos indicadores y modelos de clasificación propuestos en esta tesis deberían ser validados en otras series de pacientes, preferentemente mediante un estudio multicéntrico. También debería plantearse la combinación de modelos y/o reglas de clasificación, intentando así aumentar aún más la sensibilidad en la detección de MCA en ambos sexos mediante el ECG. En último lugar, es también aconsejable desarrollar criterios que mejoren la identificación de la MCA en mujeres, ya que los descritos en este trabajo no permiten alcanzar un rendimiento de clasificación tan elevado como el logrado en el sexo masculino.

Por otro lado, la detección de los puntos fiduciales de las ondas R y T es crucial en todos los estudios realizados en la presente tesis, al estar directamente relacionada con los rendimientos de clasificación alcanzados. En este sentido, sería interesante evaluar el rendimiento del detector de ondas empleado, mejorando su comportamiento en aquellos casos donde sea necesario.

Finalmente, en todas las ampliaciones futuras, debe tenerse en cuenta el coste computacional según las circunstancias y el dispositivo donde se vayan a utilizar los indicadores y modelos de clasificación desarrollados.

Capítulo 8

Contribuciones

8.1. Contribuciones principales de esta tesis

8.1.1. Artículos en revista

- **Jiménez-Serrano S**, Rodrigo M, Calvo CJ, Millet J, Castells F. From 12 to 1 ECG lead: multiple cardiac condition detection mixing a hybrid machine learning approach with a one-versus-rest classification strategy. *Physiological Measurement* 2022; 43:064003. DOI: 10.1088/1361-6579/ac72f5.

8.1.2. Congresos internacionales

- **Jiménez-Serrano S**, Rodrigo M, Calvo CJ, Castells F, Millet J. Multiple Cardiac Disease Detection from Minimal-Lead ECG Combining Feedforward Neural Networks with a One-vs-Rest Approach. *Computing in Cardiology 2021*; 48:1-4. DOI: 10.23919/CinC53138.2021.9662845.
- **Jiménez-Serrano S**, Yagüe-Mayans J, Simarro-Mondéjar E, Calvo CJ, Castells F, Millet J. Atrial Fibrillation Detection Using Feedforward Neural Networks and Automatically Extracted Signal Features. *Computing in Cardiology 2017*; 44:131–134. DOI: 10.23919/CinC53138.2021.9662845.
- **Jiménez-Serrano S**, Sanz-Sánchez J, Cebrián A, Igual B, Cervigón R, Millet J, Zorio E, Castells F. Study of New Criteria Based on Eigenvalue Decomposition to Assist Arrhythmogenic Cardiomyopathy Diagnosis. *Computing in Cardiology 2016*; 43:449–452.

8.1.3. Congresos nacionales

- Sanz-Sánchez J, **Jiménez-Serrano S**, Igual Muñoz B, Gimeno Blanes FJ, Gimeno Blanes JR, Millet J, Castells F, Zorio-Grima E. Heterogeneidad en la Activación Eléctrica en la Miocardiopatía Arritmogénica a partir del Vectocardiograma y el Análisis de Componentes Principales del Electrocardiograma. *Congreso de las Enfermedades Cardiovasculares (SEC). Revista Española de Cardiología* 2017; 70:764.
- Sanz-Sánchez J, **Jiménez-Serrano S**, Igual Muñoz B, Marí Alexandre J, Millet J, Martínez-Dolz L, Castells F, Zorio-Grima E. Utilidad de la Amplitud de la Señal del ECG en el Diagnóstico de Miocardiopatía Arritmogénica con Afectación del Ventrículo Izquierdo. *Congreso de las Enfermedades Cardiovasculares (SEC). Revista Española de Cardiología* 2017; 70:57.
- **Jiménez-Serrano S**, Sanz-Sánchez J, Martínez-Hinarejos CD, Igual Muñoz B, Millet J, Zorio E, Castells F. A first approach to Arrhythmogenic Cardiomyopathy detection through ECG and Hidden Markov Models. *XXXIV Congreso Anual de la Sociedad Española de Ingeniería Biomédica (CASEIB 2016)*. ISBN: 978-84-9048-531-6.
- Sanz-Sánchez J, **Jiménez-Serrano S**, Cebrián A, Igual-Muñoz B, Cervigón R, Millet J, Castells F, Zorio-Grima E. Estudio de Nuevos Criterios Diagnósticos en Miocardiopatía Arritmogénica Basados en Descomposicion de Eigenvalores. *Congreso Nacional de la Sociedad Española de Cardiología (SEC). Revista Española de Cardiología* 2016; 69:576.
- **Jiménez-Serrano S**, Sanz-Sánchez J, Calvo CJ, Domingo D, Cervigón R, Igual B, Arnau MA, Martínez L, Millet J, Zorio E, Castells F. Estudio de nuevos indicadores basados en el Vectocardiograma como marcadores diagnósticos en Cardiomiopatía Arritmogénica. *XXXIII Congreso Anual de la Sociedad Española de Ingeniería Biomédica (CASEIB 2015)*. ISBN: 978-84-608-3354-3.
- Sanz-Sánchez J, **Jiménez-Serrano S**, Domingo D, Millet J, Arnau MA, Igual B, Castells F, Zorio E. Parámetros ECG para la identificación de pacientes con Miocardiopatía Arritmogénica con afectación del ventrículo izquierdo. *Congreso de las Enfermedades Cardiovasculares de la Sociedad Española de Cardiología (SEC). Revista Española de Cardiología* 2015; 68:1054.

8.2. Contribuciones relacionadas con esta tesis

8.2.1. Artículos en revista

- Gisbert V, **Jiménez-Serrano S**, Roses-Albert E, Rodrigo M. Atrial location optimization by electrical measures for Electrocardiographic Imaging. *Computers in Biology and Medicine* 2020; 127: 104031. DOI: 10.1016/j.combiomed.2020.104031.
- Filgueiras D, Calvo C, Salvador O, Cádenas R, Ruiz J, Armada E, Rey JR, Merino JL, Peinado R, Pérez-Castellano N, Pérez-Villacastín J, Quintanilla J, **Jiménez S**, Castells F, Chorro FJ, López JL, Berenfeld O, Jalife J, López de Sá E, Millet J. Spectral analysis-based risk score enables early prediction of mortality and cerebral performance in patients undergoing therapeutic hypothermia for ventricular fibrillation and comatose status. *International Journal of Cardiology* 2015; 186:250-258. DOI: 10.1016/j.ijcard.2015.03.074.
- **Jiménez-Serrano S**, Tortajada S, García-Gómez JM. A mobile Health application to predict postpartum depression based on machine learning. *Telemedicine and e-Health* 2015; 21:567-74. DOI: 10.1089/tmj.2014.0113.

8.2.2. Congresos internacionales

- Vives-Gilabert Y, Igual B, **Jiménez-Serrano S**, Sanz-Sánchez J, Cervigón R, Cebrián A, Santabábara JM, Millet J, Zorio E, Castells F. Bayesian classification applied to strain in arrhythmogenic left-ventricle cardiomyopathy. *Computing in Cardiology 2017*. DOI: 10.22489/CinC.2017.079-088.
- Calvo CJ, Tormos A, Roses-Albert E, Zarzoso M, Arias-Mutis OJ, Cebrián A, **Jiménez-Serrano S**, Simarro-Mondéjar E, Millet J, Chorro J, Guill-Ibáñez A. Dual-sided Mapping during Global Stretch using a Custom Miniaturized Endocardial Balloon with a Multipurpose Multichannel Acquisition System for Preclinical Electrophysiological Studies. *Computing in Cardiology 2017*. DOI: 10.22489/CinC.2017.106-383.
- Yague J, **Jimenez-Serrano S**, Alonso P, Cervigón R, Calvo CJ, Castells F, Osca J, Millet J. Comparison of general purpose ECG analyzers in patients with an implanted CRT device. *Computing in Cardiology 2016*; 43:453-456.

8.2.3. Congresos nacionales

- Simarro-Mondejar E, Julià J, Castells F, **Jiménez-Serrano S**, Calvo CJ, Fontanela A, López-Gil M, Millet J. Evolución de la frecuencia dominante en fibrilación auricular accidental en el tratamiento con vernakalant. XXXIV Congreso Anual de la Sociedad Española de Ingeniería Biomédica (CASEIB 2016). ISBN: 978-84-9048-531-6.

8.3. Proyectos de investigación

Parte de este trabajo ha sido desarrollado y financiado dentro del marco del siguiente proyecto de investigación, en el que el autor ha participado:

- Universitat Politècnica de València.
Identificación de nuevos criterios electrocardiográficos diagnósticos para la miocardiopatía arritmogénica con afectación de ventrículo izquierdo - INCEMAVI (UPV-FE-14-27).
Desde el 25 de noviembre de 2014, hasta el 30 de junio de 2016.
Investigador principal: Francisco Castells Ramon.

Bibliografía

- [**Abou 2015**] Abou-Abbas L, Fersaie, Tadj C. Automatic detection of the expiratory and inspiratory phases in new born cry signals. *Electrical Biomedical Signal Processing and Control* 2015; 19: 35-43.
- [**Al-Alusi 2019**] Al-Alusi MA, Ding E, McManus DD, Lubitz SA. Wearing Your Heart on Your Sleeve: the Future of Cardiac Rhythm Monitoring. *Current Cardiology Reports* 2019; 21:158.
- [**Alcaraz 2006**] Alcaraz R, Vaya C, Cervigon R, Sanchez C, Rieta J. Wavelet sample entropy: A new approach to predict termination of atrial fibrillation. *Computing in Cardiology* 2006; 597-600.
- [**Alcaraz 2010**] Alcaraz R, Abásolo D, Hornero R, Rieta J. Optimal parameters study for sample entropy-based atrial fibrillation organization analysis. *Computer Methods and Programs in Biomedicine* 2010; 99:124-32.
- [**Alexakis 2003**] Alexakis C, Nyongesa H, Saatchi R, Harris N, Davies C, Emery C, Ireland R, Heller S. Feature extraction and classification of electrocardiogram (ECG) signals related to hypoglycemia. *Computers in Cardiology* 2003; 537-40.
- [**Allessie 1995**] Allessie M, Konings K, Wijffels M. Atrial Arrhythmias—State of the Art: Electrophysiological Mechanism of Atrial Fibrillation. *Eds. Armonk* 1995.
- [**Alpaydin 2016**] Alpaydin E. Machine Learning: The New AI. Cambridge, MA, USA. *MIT Press* 2016.
- [**Annavarapu 2016**] Annavarapu A, Kora P. ECG-based atrial fibrillation detection using different orderings of conjugate symmetric-complex hadamard transform. *International Journal Of The Cardiovascular Academy* 2016; 2:151-154.
- [**Ansari 2017**] Ansari S, Farzaneh N, Duda M, Horan K, Andersson HB, Goldberger ZD, Nallamothu BK, Najarian K. A Review of Automated Methods for Detection

- of Myocardial Ischemia and Infarction Using Electrocardiogram and Electronic Health Records. *IEEE Reviews in Biomedical Engineering* 2017;10:264-298.
- [**Antoni 2022**] Antoni L, Bruoth E, Bugata P, Bugata Jr P, Gajdos D, Horvat S, Hudak D, Kmecova V, Stana R, Stanková M, Szabari A, Vazarikova G. Automatic ECG classification and label quality in training data. *Physiological Measurement* 2022; 43:064008.
- [**Anuradha 2008**] Anuradha B, Veera V. Cardiac Arrhythmia Classification using Fuzzy Classifiers. *Journal of Theoretical and Applied Information Technology* 2008; 353-359.
- [**Aqil 2015**] Aqil M, Jbari A, Bourouhou A. Evaluation of time-frequency and wavelet analysis of ECG signals. *Third World Conference on Complex Systems (WCCS) 2015*; 1-5.
- [**Asgari 2015**] Asgari S, Mehrnia A, Moussavi M. Automatic detection of atrial fibrillation using stationary wavelet transform and support vector machine. *Computers in Biology and Medicine* 2015; 60:132-142.
- [**Aublin 2021**] Aublin P, Ammar MB, Barret M, Fix J, Oster J. Cardiac Abnormality Detection based on an Ensemble Voting of Single-Lead Classifier Predictions. *Computing in Cardiology 2021*; 43.
- [**Aublin 2022**] Aublin P, Ammar MB, Fix J, Barret M, Behar JA, Oster J. Cardiac Predict alone, decide together: cardiac abnormality detection based on single lead classifier voting. *Physiological Measurement* 2022; 43:054001.
- [**Augusto 2020**] Augusto JB, Eiros R, Nakou E, Moura-Ferreira S, Treibel TA, Captur G, Akhtar MM, Protonotarios A, Gossios TD, Savvatis K, Syrris P, Mohiddin S, Moon JC, Elliott PM, Lopes LR. Dilated cardiomyopathy and arrhythmogenic left ventricular cardiomyopathy: a comprehensive genotype-imaging phenotype study. *European Heart Journal Cardiovascular Imaging* 2020; 21:326–336.
- [**Babaeizadeh 2009**] Babaeizadeh S, Gregg R, Helfenbein E, Lindauer J, Zhou S. Improvements in atrial fibrillation detection for real-time monitoring. *Journal of Electrocardiology* 2009; 42:522-6.
- [**Barker 2022**] Barker J, Li X, Khavandi S, Koeckerling D, Mavilakandy A, Pepper C, Bountziouka V, Chen L, Kotb A, Antoun I, Mansir J, Smith-Byrne K, Schindwein FS, Dhutia H, Tyukin I, Nicolson WB, Ng GA. Machine learning in sudden cardiac death risk prediction: a systematic review. *Europace* 2022; 24:1777-1787.

- [**Bayés 2014**] Bayés de Luna A. ECGs for beginners. *Wiley-Blackwell* 2014. DOI:10.1002/9781118821350.
- [**Bazett 1920**] Bazett JC. An analysis of time relation of electrocardiograms. *Heart* 1920; 7:353-367.
- [**Bengacemi 2021**] Bengacemi H, Gharbi AH, Ravier P, Abed-Meraim Karim, Butteli O. Surface EMG signal segmentation based on HMM modelling: Application on Parkinson's disease. *ENP Engineering Science Journal* 2021; 1:63-74.
- [**Bhattacharyya 2020**] Bhattacharyya S, Goswami DP, Sengupta A. Spatial velocity of the dynamic vectorcardiographic loop provides crucial insight in ventricular dysfunction. *Medical Hypotheses* 2020; 135:109484.
- [**Bickerton 2019**] Bickerton M, Pooler A. Misplaced ECG electrodes and the need for continuing training. *British Journal of Cardiac Nursing* 2019; 14:123-32.
- [**Bin 2017**] Bin G, Shao M, Bin G, Huang J, Zheng D, Wu S. Detection of Atrial Fibrillation Using Decision Tree Ensemble. *Computing in Cardiology* 2017; 44.
- [**Birger 2021**] Birger M, Kaldjian AS, Roth GA, Moran AE, Dieleman JL, Bellows BK. Spending on Cardiovascular Disease and Cardiovascular Risk Factors in the United States: 1996 to 2016. *Circulation* 2021;144:271-282.
- [**Bishop 2006**] Bishop C. Pattern Recognition and Machine Learning. *Springer* 2006.
- [**Blik 2018**] Blik EC. ST elevation: Differential diagnosis and caveats. A comprehensive review to help distinguish ST elevation myocardial infarction from nonischemic etiologies of ST elevation. *Turkish Journal of Emergency Medicine*. 2018; 18:1-10.
- [**Bollmann 1998**] Bollmann A, Kanuru N, McTeague K, Walter P, DeLurgio D, Langberg J. Frequency analysis of human atrial fibrillation using the surface electrocardiogram and its response to ibutilide. *American Journal of Cardiology* 1998; 81:1439-45.
- [**Bollman 2002a**] Bollmann A, Biniyas K, Toepffer I, Molling J, Geller C, Klein H. Importance of left atrial diameter and atrial fibrillatory frequency for conversion of persistent atrial fibrillation with oral flecainide. *American Journal of Cardiology* 2002; 90:1011-1014.
- [**Bollman 2002b**] Bollmann A, Sonne K, Esperer H, Toepffer I, Klein H. Patients with persistent atrial fibrillation taking oral verapamil exhibit a lower atrial frequency on the ECG. *Annals of Noninvasive Electrocardiology* 2002; 7:92-97.

- [**Bollman 2006**] Bollmann A, Husser D, Mainardi L, Lombardi F, Langley P, Murray A, Rieta J, Millet J, Olsson S, Stridh M, Sörnmo L. Analysis of surface electrocardiograms in atrial fibrillation: Techniques, research, and clinical applications. *Europace* 2006; 8:911-26.
- [**Bonizzi 2010**] Bonizzi P, Guillem MS, Climent AM, Millet J, Zarzoso V, Castells F, Meste O. Noninvasive assessment of the complexity and stationarity of the atrial wavefront patterns during atrial fibrillation. *IEEE Transactions on Biomedical Engineering* 2010; 57:2147-2157.
- [**Bortolan 2001**] Bortolan G, Christov I. Myocardial infarction and ischemia characterization from T-loop morphology in VCG. *Computing in Cardiology* 2001; 28:633–636.
- [**Bortolan 2002**] Bortolan G, Bressan M, Christov I. Longitudinal modifications of T-loop morphology. *Computing in Cardiology* 2002; 29:685–688.
- [**Bortolan 2022**] Bortolan G, Christov I, Simova I. Modifications in Electrocardiographic and Vectordardiographic Morphological Parameters in Elderly Males as Result of Cardiovascular Diseases and Diabetes Mellitus. *Diagnostics* 2022; 12:2911.
- [**Bousseljot 1995**] Bousseljot R, Kreiseler D, Schnabel A. Nutzung der EKG-Signaldatenbank CARDIODAT der PTB über das Internet. *Biomedizinische Technik/Biomed Eng* 1995; 40: 317–18.
- [**Bumgarner 2018**] Bumgarner JM, Lambert CT, Hussein AA, Cantillon DJ, Baranowski B, Wolski K, Lindsay BD, Wazni OM, Tarakji KG. Smartwatch algorithm for automated detection of atrial fibrillation. *Journal of the American College of Cardiology* 2018; 71:2381-2388.
- [**Carrara 2015**] Carrara M, Carozzi L, Moss T, Pasquale M, Cerutti S, Ferrario M, Lake D, Moorman J. Heart rate dynamics distinguish among atrial fibrillation, normal sinus rhythm and sinus rhythm with frequent ectopy. *Physiological Measurement* 2015; 36:1873-88.
- [**Calkins 2017**] Calkins H, Corrado D, Marcus F. Risk Stratification in Arrhythmogenic Right Ventricular Cardiomyopathy. *Circulation* 2017; 136:2068–208.
- [**Castells 2005a**] Castells F, Mora C, Rieta J, Moratal-Pérez D, Millet J. Estimation of atrial fibrillatory wave from single-lead atrial fibrillation electrocardiograms using principal component analysis concepts. *Medical & Biological Engineering & Computing* 2005; 43:557-60.

- [**Castells 2005b**] Castells F, Rieta J, Millet J, Zarzoso V. Spatiotemporal Blind Source Separation Approach to Atrial Activity Estimation in Atrial Tachyarrhythmias. *IEEE Transactions on Biomedical Engineering* 2005; 52:258-67.
- [**Castells 2007**] Castells F, Laguna P, Sörnmo L, Bollmann A, Millet J. Principal Component Analysis in ECG Signal Processing. *EURASIP Journal on Advances in Signal Processing* 2007; 1:98-119.
- [**Castells 2011**] Castells F, Meste O, Quesada A, Guillem MS, Climent AM, Millet J. Characterization of typical and atypical Atrial Flutter Loops from the Vectorcardiogram. *33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society 2011*; 4976-9.
- [**Cerutti 2008**] Cerutti S, Mainardi L, Sörnmo L. Understanding atrial fibrillation: the signal processing contribution. *Morgan and Claypool Publishers* 2008.
- [**Chang 2012**] Chang P, et al. Myocardial infarction classification with multi-lead ECG using hidden Markov models and Gaussian mixture models. *Applied Soft Computing Journal* 2012; 12:3165-75.
- [**Chazal 2004**] Chazal P, O'Dwyer M, Reilly RB. Automatic classification of heartbeats using ECG morphology and heartbeat interval features. *IEEE Transactions on Biomedical Engineering* 2004; 51:1196-1206.
- [**Chen 2016**] Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'16)*. 2016; 785–794.
- [**Chen 2018**] Chen Y, Wang X, Jung Y, Abedi V, Zand R, Bikak M, Adibuzzaman M. Classification of short single-lead electrocardiograms (ECGs) for atrial fibrillation detection using piecewise linear spline and xgboost. *Physiological Measurement* 2018; 39 104006.
- [**Chow 2012**] Chow GV, Marine JE, Fleg JL. Epidemiology of Arrhythmias and Conduction Disorders in Older Adults. *Clinics in Geriatric Medicine* 2012; 28:539–553.
- [**Christov 2006**] Christov I, Gómez-Herrero G, Krasteva V, Jekova I, Gotchev A, Egiazarian K. Comparative study of morphological and time-frequency ECG descriptors for heartbeat classification. *Medical Engineering & Physics* 2006; 28:876-887.

- [**Chugh 2014**] Chugh S, Havmoeller R, Narayanan K, Singh D, Rienstra M, Benjamin E and Murray C. Worldwide Epidemiology of Atrial Fibrillation: A Global Burden of Disease 2010 Study. *Circulation* 2014; 129:837-47.
- [**Ciaccio 2012**] Ciaccio E, Biviano A, Whang W, Garan H. A new LMS algorithm for analysis of atrial fibrillation signals. *BioMedICAL Engineering Online* 2012; 11-15.
- [**Clifford 2017**] Clifford G, Liu C, Moody B, Silva I, Li Q, Johnson A, Mark R. AF Classification from a Short Single Lead ECG Recording: the PhysioNet Computing in Cardiology Challenge 2017. *Computing in Cardiology 2017*; 44.
- [**Cipriani 2020**] Cipriani A, Bauce B, Lazzari M, Rigato I, Bariani R, Meneghin S, Pilichou K, Motta R, Aliberti C, Thiene G, McKenna WJ, Zorzi A, Iliceto S, Basso C, Perazzolo M, Corrado D. Arrhythmogenic right ventricular cardiomyopathy: characterization of left ventricular phenotype and differential diagnosis with dilated cardiomyopathy. *Journal of the American Heart Association* 2020; 9:e014628.
- [**Colloca 2013a**] Colloca R, Minardi L, Clifford G. Implementation and testing of atrial fibrillation detectors for a mobile phone application. *Politecnico di Milano & University of Oxford* 2013.
- [**Colloca 2013b**] Colloca R, Johnson A, Mainardi L, Clifford G. A support vector machine approach for reliable detection of atrial fibrillation events. *Computing in Cardiology 2013*; 1047-1050.
- [**Corrado 2011**] Corrado D, Basso C, Pilichou K, Thiene G. Molecular biology and the clinical management of arrhythmogenic right ventricular cardiomyopathy/dysplasia. *Heart* 2011; 97:530-9.
- [**Corrado 2017**] Corrado D, Basso C, Judge DP. Arrhythmogenic cardiomyopathy. *Circulation Research* 2017; 121:784-802.
- [**Corrado 2020**] Corrado D, Perazzolo Marra M, Zorzi A, Beffagna G, Cipriani A, Lazzari MD, Migliore F, Pilichou K, Rampazzo A, Rigato I, et al. Diagnosis of arrhythmogenic cardiomyopathy: the Padua criteria. *International Journal of Cardiology* 2020;319:106-114.
- [**Corrado 2021**] Corrado D, Zorzi A, Cipriani A, Bauce B, Bariani R, Beffagna G, Lazzari M, Migliore F, Pilichou K. Evolving Diagnostic Criteria for Arrhythmogenic Cardiomyopathy. *Journal of the American Heart Association* 2021; 10:e021987.

- [Cox 1958] Cox DR. The Regression Analysis of Binary Sequences. *Journal of the Royal Statistical Society* 1958; 20:215-242.
- [Cristianini 2000] Cristianini N, Shawe-Taylor J. An introduction to Support Vector Machines and other kernel-based learning methods. *Cambridge University Press* 2000.
- [Dai 2021] Dai H, Hwang HG, Tseng VS. Convolutional neural network based automatic screening tool for cardiovascular diseases using different intervals of ECG signals. *Computer Methods and Programs in Biomedicine* 2021; 203:106035.
- [Daliento 1995] Daliento L, Turrini P, Nava A, Rizzoli G, Angelini A, Buja G, Scognamiglio R, Thiene G. Arrhythmogenic right ventricular cardiomyopathy in young versus adult patients: similarities and differences. *Journal of the American College of Cardiology* 1995; 25:655–664.
- [Datta 2017] Datta S, Puri C, Mukherjee A, Banerjee R, Choudhury AD, Singh R, Ukil A, Bandyopadhyay S, Pal A, Khandelwal S. Identifying Normal, AF and other Abnormal ECG Rhythms using a Cascaded Binary Classifier. *Computing in Cardiology* 2017; 44.
- [Dedeoglu 2022] Dedeoglu S, Dede E, Oztunc F, Gedikbasi A, Yesil G, Dedeoglu R. Mutation identification and prediction for severe cardiomyopathy in Alström syndrome, and review of the literature for cardiomyopathy. *Orphanet journal of rare diseases* 2022; 17:359.
- [DeMazumder 2013] DeMazumder D, Lake D, Cheng A, Moss T, Guallar E, Weiss R, Jones S, Tomaselli G, Moorman J. Dynamic analysis of cardiac rhythms for discriminating atrial fibrillation from lethal ventricular arrhythmias. *Circulation: Arrhythmia and Electrophysiology* 2013; 6:555-561.
- [Denysyuk 2023] Denysyuk HV, Pinto RJ, Silva PM, Duarte RP, Marinho FA, Pimenta L, Gouveia AJ, Gonçalves NJ, Coelho PJ, Zdravevski E, Lameski P, Leithardt V, Garcia NM, Pires IM. Algorithms for automated diagnosis of cardiovascular diseases based on ECG data: A comprehensive systematic review. *Heliyon* 2023; 9:e13601.
- [Deshpande 2014] Deshpande A, Birnbaum Y. ST-segment elevation: Distinguishing ST elevation myocardial infarction from ST elevation secondary to nonischemic etiologies. *World journal of cardiology* 2014; 6:1067–1079.
- [Dower 1984] Dower GE. The ECGD: A derivation of the ECG from VCG leads. *Journal of Electrocardiology* 1984; 17:189–191.

- [**Du 2014**] Du X, Rao N, Qian M, Liu D, Li J, Feng W, Yin L, Chen X. A novel method for real-time atrial fibrillation detection in electrocardiograms using multiple parameters. *Annals of Noninvasive Electrocardiology* 2014; 19:217-25.
- [**Duncker 2021**] Duncker D, Ding WY, Etheridge S, Noseworthy PA, Veltmann C, Yao X, Bunch TJ, Gupta D. Smart Wearables for Cardiac Monitoring-Real-World Use beyond Atrial Fibrillation. *Sensors (Basel)* 2021; 21:2539.
- [**Dunn 2018**] Dunn J, Runge R, Snyder M. Wearables and the medical revolution. *Personalized Medicine* 2018; 15:429-448.
- [**Edenbrandt 1988**] Edenbrandt L, Pahlm O. Vectorcardiogram synthesized from a 12-lead ECG: Superiority of the inverse Dower matrix. *Journal of Electrocardiology* 1988; 21:361-7.
- [**Elola 2018**] Elola A., Aramendi E., Irusta U. et al. ECG-based pulse detection during cardiac arrest using random forest classifier. *Medical and Biological Engineering and Computing* 2018; 57:453-462.
- [**Esperer 2008**] Esperer H, Esperer C, Cohen R. Cardiac arrhythmias imprint specific signatures on Lorenz plots. *Annals of Noninvasive Electrocardiology* 2008; 13:44-6.
- [**Fan 2018**] Fan X, Yao Q, Cai Y, Miao F, Sun F, Li Y. Multiscaled Fusion of Deep Convolutional Neural Networks for Screening Atrial Fibrillation From Single Lead Short ECG Recordings. *IEEE Journal of Biomedical and Health Informatics* 2018; 22:1744-1753.
- [**Feeny 2020**] Feeny AK, Rickard J, Trulock KM, Patel D, Toro S, Moennich LA, Varma N, Niebauer MJ, Gorodeski EZ, Grimm RA, Barnard J, Madabhushi A, Chung MK. Machine Learning of 12-Lead QRS Waveforms to Identify Cardiac Resynchronization Therapy Patients With Differential Outcomes. *Circulation: Arrhythmia and Electrophysiology* 2020;13:e008210.
- [**Fridericia 2003**] Fridericia LS. The Duration of Systole in an Electrocardiogram in Normal Humans and in Patients with Heart Disease. *Annals of Noninvasive Electrocardiology* 2003; 8:343-351.
- [**Friesen 1990**] Friesen GM, Jannett TC, Jadallah MA, Yates SL, Quint SR, Nagle HT. A comparison of the noise sensitivity of nine QRS detection algorithms. *IEEE Transactions on Biomedical Engineering* 1990; 37:85-98.
- [**Fu 2015**] Fu DG. Cardiac Arrhythmias: Diagnosis, Symptoms, and Treatments. *Cell Biochemistry and Biophysics* 2015; 73:291-296.

- [Fuster 2001] Fuster V, Ryden L, Asinger R, Cannom D, Crijs H, Frye R. ACC/AHA/ESC guidelines for the management of patients with atrial fibrillation: executive summary. *Journal of the American College of Cardiology* 2001; 38:1231-65.
- [Garabelli 2017] Garabelli P, Stavrakis S, Po S. Smartphone-based arrhythmia monitoring. *Current Opinion in Cardiology* 2017; 32:53-57.
- [García 2016] García M, Ródenas J, Alcaraz R, Rieta J. Application of the relative wavelet energy to heart rate independent detection of atrial fibrillation. *Computer Methods and Programs in Biomedicine* 2016; 131:157-68.
- [Gaziano 2010] Gaziano TA, Bitton A, Anand S, Abrahams-Gessel S and Murphy A. Growing epidemic of coronary heart disease in low and middle income countries. *Current Problems in Cardiology* 2010; 35:72–115.
- [Gaztañaga 2012] Gaztañaga L, Marchlinski F, Betensky B. Mecanismos de las arritmias cardiacas. *Revista Española de Cardiología* 2012; 65:174-85.
- [Georgiou 2018] Georgiou K, Larentzakis AV, Khamis NN, Alsuhaibani GI, Alaska YA, Giallafos EJ. Can wearable devices accurately measure heart rate variability? A systematic review. *Folia Medica* 2018; 60:7-20.
- [Gheorghe 2018] Gheorghe A, Griffiths U, Murphy A, Legido-Quigley H, Lamptey P, Perel P. The economic burden of cardiovascular disease and hypertension in low- and middle-income countries: a systematic review. *BMC Public Health* 2018; 18:975.
- [Gondra 2008] Gondra I. Applying machine learning to software fault-proneness prediction. *The Journal of Systems and Software* 2008; 81:186-95.
- [Giovanni 2021] Giovanni B, Christov I, Simova I. Potential of Rule-Based Methods and Deep Learning Architectures for ECG Diagnostics. *Diagnostics* 2021; 11:1678.
- [Guyton 2006] Guyton AC, Hall J. Textbook of Medical Physiology, 11th edition. *Elsevier Saunders* 2006.
- [Hagan 2021] Hagan R, Gillan CJ, Mallett F. Comparison of machine learning methods for the classification of cardiovascular disease. *Informatics in Medicine Unlocked* 2021; 24 100606.
- [Harmon 2023] Harmon DM, Noseworthy PA, Yao X. The Digitization and Decentralization of Clinical Trials. *Mayo Clinic Proceedings* 2023; 18:S0025-6196(22)00560-2.

- [**Hartmann 2019**] Hartmann M, Hashmi US, Imran A. Edge computing in smart health care systems: Review challenges and research directions. *Transactions on Emerging Telecommunications Technologies* 2019.
- [**Hasna 2017**] Hasna OL, Potolea R. Robust Feature Extraction from Noisy ECG for Atrial Fibrillation Detection. *Computing in Cardiology* 2017; 44.
- [**Haugaa 2016**] Haugaa KH, Haland TF, Leren IS, Saberniak J, Edvardsen T. Arrhythmogenic right ventricular cardiomyopathy, clinical manifestations, and diagnosis. *Europace*. 2016; 18:965–72.
- [**Hermans 2022**] Hermans ANL, Gawalko M, Dohmen L, van der Velden RMJ, Betz K, Duncker D, Verhaert DVM, Heidbuchel H, Svennberg E, Neubeck L, Eckstein J, Lane DA, Lip GYH, Crijns HJGM, Sanders P, Hendriks JM, Pluymaekers NAHA, Linz D. Mobile health solutions for atrial fibrillation detection and management: a systematic review. *Clinical Research in Cardiology* 2022; 111:479-491.
- [**Hindricks 2021**] Hindricks G, Potpara T, Dagres N, Arbelo E, Bax JJ, Blomström-Lundqvist C, et. al. 2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS): The Task Force for the diagnosis and management of atrial fibrillation of the European Society of Cardiology (ESC). *European Heart Journal* 2021; 42:373-498.
- [**Hong 2017**] Hong H, Wu M, Zhou Y, Wang Q, Shang J, Li H, Xie J. ENCASE: an ENsemble CIASsifiEr for ECG Classification Using Expert Features and Deep Neural Networks. *Computing in Cardiology* 2017; 44.
- [**Hoorntje 2017**] Hoorntje ET, Te Rijdt WP, James CA, Pilichou K, Basso C, Judge DP, Bezzina CR, Tintelen JP. Arrhythmogenic cardiomyopathy: pathology, genetics, and concepts in pathogenesis. *Cardiovascular Research* 2017; 113:1521–1531.
- [**Hosmer 2013**] Hosmer DW Jr, Lemeshow S, Sturdivant RX. Applied Logistic Regression. *Wiley* 2013.
- [**Hsu 2002**] Hsu C y Lin CJ. A comparison of methods for multiclass support vector machines. *IEEE transactions on Neural Networks* 2002; 13:415-25.
- [**Hu 2001**] Hu Y, Hwang J. Handbook of neural networks signal processing. *CRC Press* 2001.

- [**Hua 2021**] Hua Z, Chengyu L, Zhimin Z, Yujie X, Xinwen L, Ruiqing D, Yu H, Ling X, Feng L. Recurrence Plot-Based Approach for Cardiac Arrhythmia Classification Using Inception-ResNet-v2. *Frontiers in Physiology* 2021; 12.
- [**Huang 2011**] Huang C, Ye S, Chen H, Li D, He F, Tu Y. A novel method for detection of the transition between atrial fibrillation and sinus rhythm. *IEEE Transactions on Biomedical Engineering* 2011; 58:1113-9.
- [**Hume 2012**] Hume J, Grant A. Agents Used in Cardiac Arrhythmias. Basic & Clinical Pharmacology, 13th edition. *McGraw-Hill Medical* 2012.
- [**Hussein 2018**] Hussein AF, Kumar NA, Burbano-Fernandez M, Ramírez-González G, Abdulhay E, De Albuquerque VH. An automated remote cloud-based heart rate variability monitoring system. *IEEE Access* 2018; 6:77055-77064.
- [**Isakadze 2020**] Isakadze N, Martin SS. How useful is the smartwatch ECG? *Trends in Cardiovascular Medicine* 2020; 30:442.
- [**Jalife 2004**] Jalife J. Los rotores y la fibrilación. *Archivos de Cardiología de México* 2004; 74:290-2.
- [**Japkowicz 2002**] Japkowicz N, Stephen S. The class imbalance problem: a systematic study. *Intelligent data analysis Journal* 2002; 6:429-49.
- [**Jiménez 2015**] Jiménez-Serrano S, Sanz-Sánchez J, Calvo CJ, Domingo D, Cervigón R, Igual B, Arnau MA, Martínez L, Millet J, Zorio E, Castells F. Estudio de nuevos indicadores basados en el Vectocardiograma como marcadores diagnósticos en Cardiomiopatía Arritmogénica. *XXXIII Congreso Anual de la Sociedad Española de Ingeniería Biomédica (CASEIB 2015)*. ISBN: 978-84-608-3354-3.
- [**Jiménez 2016a**] Jiménez-Serrano S, Sanz-Sánchez J, Cebrián A, Igual B, Cervigón R, Millet J, Zorio E, Castells F. Study of New Criteria Based on Eigenvalue Decomposition to Assist Arrhythmogenic Cardiomyopathy Diagnosis. *Computing in Cardiology* 2016; 43:449-452.
- [**Jiménez 2016b**] Jiménez-Serrano S, Sanz-Sánchez J, Martínez-Hinarejos CD, Igual Muñoz B, Millet J, Zorio E, Castells F. A first approach to Arrhythmogenic Cardiomyopathy detection through ECG and Hidden Markov Models. *XXXIV Congreso Anual de la Sociedad Española de Ingeniería Biomédica (CASEIB 2016)*. ISBN: 978-84-9048-531-6.
- [**Jiménez 2017a**] Jiménez-Serrano S, Yagüe-Mayans J, Simarro-Mondéjar E, Calvo CJ, Castells F, Millet J. Atrial Fibrillation Detection Using Feedforward Neural

- Networks and Automatically Extracted Signal Features. *Computing in Cardiology 2017*; 44:131–134.
- [**Jiménez 2017b**] Sanz-Sánchez J, Jiménez-Serrano S*, Igual Muñoz B, Marí Alexandre J, Millet J, Martínez-Dolz L, Castells F, Zorio-Grima E. Utilidad de la Amplitud de la Señal del ECG en el Diagnóstico de Miocardiopatía Arritmogénica con Afectación del Ventrículo Izquierdo. *Congreso de las Enfermedades Cardiovasculares (SEC). Revista Española de Cardiología 2017*; 70:57.
- [**Jiménez 2021**] Jiménez-Serrano S, Rodrigo M, Calvo CJ, Castells F, Millet J. Multiple Cardiac Disease Detection from Minimal-Lead ECG Combining Feedforward Neural Networks with a One-vs-Rest Approach. *Computing in Cardiology 2021*; 48:1-4.
- [**Jiménez 2022**] Jiménez-Serrano S, Rodrigo M, Calvo CJ, Millet J, Castells F. From 12 to 1 ECG lead: multiple cardiac condition detection mixing a hybrid machine learning approach with a one-versus-rest classification strategy. *Physiological Measurement 2022*; 43:064003.
- [**Jo 2021**] Jo YY, Kwon JM, Jeon KH, Cho YH, Shin JH, Lee YJ, Jung MS, Ban JH, Kim KH, Lee SY, Park J, Oh BH. Detection and classification of arrhythmia using an explainable deep learning model. *Journal of Electrocardiology 2021*; 67:124-132.
- [**Joliffe 2002**] Joliffe IT. Principal Component Analysis. *Springer 2002*.
- [**Kallert 1999**] Kallert T, Couderc JP, Voss A, Zareba W. Semi-automatic method quantifying T wave loop morphology: Relevance for assessment of heterogeneous repolarization. *Computing in Cardiology 1999*; 26:153–156.
- [**Kang 2022**] Kang J, Weng H. A study on several critical problems on arrhythmia detection using varying-dimensional electrocardiography. *Physiological Measurement 2022*; 43:064007.
- [**Khalifa 2021**] Khalifa Y, Mandic D, Sejdic E. A review of Hidden Markov models and Recurrent Neural Networks for event detection and localization in biomedical signals. *Information Fusion 2021*; 69:52-72.
- [**Kligfield 2002**] Kligfield P. The centennial of the Einthoven electrocardiogram. *Journal of Electrocardiology 2002*; 35:123–129.
- [**Kohavi 1997**] Kohavi R, John GH. Wrappers for feature subset selection. *Artificial intelligence 1997*; 97:273-324.

- [**Kora 2016**] Kora P, Krishna KS. ECG based heart arrhythmia detection using wavelet coherence and bat algorithm. *Sensing and Imaging* 2016; 17.
- [**Kora 2017**] Kora P, Annavarapu A, Yadlapalli P, Krishna K, Somalaraju V. ECG based atrial fibrillation detection using sequency ordered complex hadamard transform and hybrid firefly algorithm. *Engineering Science and Technology, an International Journal* 2017; 20:1084-1091.
- [**Krasteva 2021**] Krasteva V, Christov I, Naydenov S, Stoyanov T, Jekova I. Application of Dense Neural Networks for Detection of Atrial Fibrillation and Ranking of Augmented ECG Feature Set. *Sensors* 2021; 20: 6848.
- [**Kostka 2007**] Kostka P, Tkacz E. Feature extraction and selection algorithms in biomedical data classifiers based on time-frequency and principle component analysis. *11th Mediterranean Conference on Medical and Biomedical Engineering and Computing 2007. IFMBE Proceedings*, 16.
- [**Kumari 2017**] Kumari P, Mathew L, Syal P. Increasing Trend of Wearables and Multimodal Interface for Human Activity Monitoring: A review. *Biosensors and Bioelectronics* 2017; 90:298-307.
- [**Kumar 2022**] Kumar A, Siddharth V, Singh SI, Narang R. Cost analysis of treating cardiovascular diseases in a super-specialty hospital. *PLoS One*. 2022 Jan 5;17(1):e0262190.
- [**Lake 2011**] Lake D, Moorman J. Accurate estimation of entropy in very short physiological time series: the problem of atrial fibrillation detection in implanted ventricular devices. *American Journal of Physiology-Heart and Circulatory Physiology* 2011; 300:319-25.
- [**Ladavich 2015**] Ladavich S, Ghoraani B. Rate-independent detection of atrial fibrillation by statistical modeling of atrial activity. *Biomedical Signal Processing and Control* 2015; 18:274-81.
- [**Lastre 2019**] Lastre-Domínguez C, Shmaliy YS, Ibarra-Manzano O, Munoz-Minjares J, Morales-Mendoza LJ. ECG Signal Denoising and Features Extraction Using Unbiased FIR Smoothing. *Biomed Research International* 2019; 20:2608547.
- [**Lawin 2022**] Lawin D, Kuhn S, Schulze Lammers S, Lawrenz T, Stellbrink C. Use of digital health applications for the detection of atrial fibrillation. *Herzschrittmacherther Elektrophysiol* 2022; 33:373-379.

- [**Li 2016**] Li Q, Liu C, Oster J, Clifford G. Signal processing and feature selection preprocessing for classification in noisy healthcare data. *Machine Learning for Healthcare Technologies* 2016; 33-58.
- [**Liang 2014**] Liang W, et al. A Novel Approach to ECG Classification Based upon Two-Layered HMMs in Body Sensor Networks. *Sensors* 2014; 14:5994-6011.
- [**Lim 2016**] Lim HW, Hau YW, Lim CW, Othman MA. Artificial intelligence classification methods of atrial fibrillation with implementation technology. *Computer Assisted Surgery* 2016; 21:154-161.
- [**Linker 2016**] Linker D. Accurate, Automated Detection of Atrial Fibrillation in Ambulatory Recordings. *Cardiovascular Engineering and Technology* 2016; 7:182-9.
- [**Lip 2016**] Lip G, Fauchier L, Freedman S, Gelder I, Natale A, Gianni C, Nattel S, Potpara T, Rienstra M, Tse H, Lane D. Atrial fibrillation. *Nature Reviews Disease Primers* 2016; 31.
- [**Liu 2018**] Liu F, et al. An Open Access Database for Evaluating the Algorithms of Electrocardiogram Rhythm and Morphology Abnormality Detection. *Journal of Medical Imaging and Health Informatics* 2018; 8:1368–1373.
- [**Luz 2013**] Luz EJ, Nunes TM, Albuquerque VH, Papa JP, Menotti D. ECG arrhythmia classification based on optimum-path forest. *Expert Systems with Applications* 2013; 40:3561-3573.
- [**Macfarlane 1994**] Macfarlane PW, McLaughlin SC, Devine B, et al. Effects of age, sex, and race on ECG interval measurements. *Journal of electrocardiology* 1994; 27:14-19.
- [**Mahmoodabadi 2005**] Mahmoodabadi S, Ahmadian A, Abolhasani M. ECG feature extraction using Daubechies wavelets. *Proceedings of the 5th IASTED International Conference on Visualization, Imaging and Image Processing* 2005; 343–348.
- [**Mahmoud 2006**] Mahmoud SS, Hussain ZM, Cosic I, Fang Q. Time-frequency analysis of normal and abnormal biological signals. *Biomedical Signal Processing and Control* 2006; 1:33-43.
- [**Mainardi 2008**] Mainardi L, Sörnmo L, Cerutti S. Understanding Atrial Fibrillation: the Signal Processing Contribution. *Morgan and Claypool Publishers* 2008.

- [**Marcus 1982**] Marcus FI, Fontaine GH, Guiraudon G, Frank R, Laurenceau JL, Malerque C, Grosgeat Y. Right ventricular dysplasia: a report of 24 adult cases. *Circulation* 1982; 65:384-98.
- [**Marcus 2010**] Marcus FI, McKenna WJ, Sherrill D, Basso C, Baucé B, Bluemke DA, Calkins H, Corrado D, Cox MG, Daubert JP, Fontaine G, Gear K, Hauer R, Nava A, Picard MH, Protonotarios N, Saffitz JE, Sanborn DM, Steinberg JS, Tandri H, Thiene G, Towbin JA, Tsatsopoulou A, Wichter T, Zareba W. Diagnosis of Arrhythmogenic Right Ventricular Cardiomyopathy/Dysplasia: proposed modification of the task force criteria. *Circulation* 2010; 121: 1533-1541.
- [**Marinho 2019**] Marinho LB, Nascimento N, Souza JW, Gurgel MV, Rebouças PP Filho, Albuquerque VH. A novel electrocardiogram feature extraction approach for cardiac arrhythmia classification. *Future Generation Computer Systems* 2019; 97:564-577.
- [**Markman 2019**] Markman TM, Nazarian S. Treatment of ventricular arrhythmias: What's New? *Trends in Cardiovascular Medicine* 2019; 29:249-261.
- [**Martínez 2004**] Martínez JP, Almeida R, Olmos S, Rocha AP, Laguna P. A wavelet-based ecg delineator: evaluation on standard databases. *IEEE Transactions on Biomedical Engineering* 2004; 51:570–81.
- [**Mateo 2013**] Mateo J, Rieta J. Radial basis function neural networks applied to efficient QRST cancellation in atrial fibrillation. *Computers in Biology and Medicine* 2013; 43:154-63.
- [**McKenna 2017**] McKenna WJ, Maron BJ, Thiene G. Classification, Epidemiology, and Global Burden of Cardiomyopathies. *Circulation Research* 2017; 121:722-730.
- [**Medeiros 2007**] Medeiros-Domingo A, Iturralde-Torres P, Ackermann MJ. Clinical and Genetic Characteristics of Long QT Syndrome. *Revista Española de Cardiología* 2007; 60:739-752.
- [**Mietus 2002**] Mietus JE, Peng C-K, Henry I, Goldsmith RL, Goldberger AL. The pNNx Files: Re-examining a Widely Used Heart Rate Variability Measure. *Heart* 2002; 88:378-380.
- [**Mihandoost 2022**] Mihandoost S, Sörnmo L, Doyen M, Oster J. A comparative study of the performance of methods for f-wave extraction. *Physiological Measurement* 2022; 43: 105006.

- [**Militaru 2019**] Militaru S, Adam R, Dorobantu L, Ferrazzi P, Iascone M, Radoi V, Ismail G, Popescu BA, Jurcut R. Rare presentation and wide intrafamilial variability of Fabry disease: A case report and review of the literature. *The Anatolian Journal of Cardiology* 2019; 22:154-158.
- [**Minami 1999**] Minami K, Nakajima H, Toyoshima T. Real-time discrimination of ventricular tachyarrhythmia with Fourier-transform neural network. *IEEE Transactions on Biomedical Engineering* 1999; 46:179–85.
- [**Mitchell 1997**] Mitchell T. Machine Learning. *McGraw Hill* 1997.
- [**Montes 2013**] Montes-Santiago J, Rodil V, Formiga F, Cepeda J and Urrutia A. Características y costes de los pacientes ingresados por arritmias cardiacas en España. *Revista Clínica Española* 2013; 213:235-9.
- [**Moss 2010**] Moss AJ. Gender Differences in ECG Parameters and Their Clinical Implications. *Annals of Noninvasive Electrocardiology* 2010; 15:1-2.
- [**Murphy 2012**] Murphy KP. Machine Learning: A Probabilistic Perspective. Cambridge, MA, USA. *MIT Press* 2012.
- [**Muser 2019**] Muser D, Santangeli P, Castro SA, Casado-Arroyo R, Maeda S, Benhayon DA, Liuba I, Liang JJ, Sadek MM, Chahal A, Magnani S, Pieroni M, Santarossa E, Desjardins B, Dixit S, Garcia FC, Callans DJ, Frankel DS, Alavi A, Marchlinski FE, Selvanayagam JB, Nucifora G. Risk stratification of patients with apparently idiopathic premature ventricular contractions. A Multicenter International CMR Registry. *JACC: Clinical Electrophysiology*. 2019; 6:722-735.
- [**Nabil 2015**] Nabil D, Reguig FB. Ectopic beats detection and correction methods: A review. *Biomedical Signal Processing and Control* 2015; 18:228-244.
- [**Naccarelli 2009**] Naccarelli G, Varker H, Lin J and Schulman K. Increasing prevalence of atrial fibrillation and flutter in the United States. *American Journal of Cardiology* 2009; 104:1534-39.
- [**Nascimento 2019**] Nascimento NM, Marinho LB, Peixoto SA, Vale Madeiro JP, Albuquerque VH, Rebouças PP. Heart arrhythmia classification based on statistical moments and structural co-occurrence. *Circuits, Systems, and Signal Processing*. 2019; 1:631-650.
- [**Nava 1988**] Nava A, Thiene G, Canciani B, Scognamiglio R, Daliento L, Buja G, Martini B, Stritoni P, Fasoli G. Familial occurrence of right ventricular dysplasia: a study involving nine families. *Journal of the American College of Cardiology*, 1988; 1222–1228.

- [**Nejedly 2022**] Nejedly P, Ivora A, Viscor I, Koscova Z, Smisek R, Jurak P, Plesinger F. Classification of ECG using ensemble of residual CNNs with or without attention mechanism. *Physiological Measurement* 2022; 43:044001.
- [**Nguyen 2021**] Nguyen QH, BP Nguyen BP, Nguyen TB, Do T, Mbinta JF, Simpson CR. Stacking segment-based CNN with SVM for recognition of atrial fibrillation from single-lead ECG recordings. *Biomedical Signal Processing and Control* 2021; 68:102672.
- [**Oster 2015**] Oster J, Clifford GD. Impact of the presence of noise on RR interval-based atrial fibrillation detection. *Journal of Electrocardiology* 2015; 48:947-51.
- [**Palm 2012**] Palm R. Prediction as a candidate for learning deep hierarchical models of data. *Technical University of Denmark* 2012.
- [**Pan 1985**] Pan J and Tompkins WJ. A real-time QRS detection algorithm. *IEEE Transactions on Biomedical Engineering* 1985; 32:230–6.
- [**Pandey 2022**] Pandey SK, Kumar G, Shukla S, Kumar A, Singh KU, Mahato S. Automatic Detection of Atrial Fibrillation from ECG Signal Using Hybrid Deep Learning Techniques. *Journal of Sensors* 2022; 6732150.
- [**Park 2009**] Park J, Lee S, Jeon M. Atrial fibrillation detection by heart rate variability in Poincare plot. *BioMedical Engineering OnLine* 2009; 8.
- [**Parvaneh 2019**] Parvaneh S, Rubin J, Babaeizadeh S, Xu-Wilson M. Cardiac arrhythmia detection using deep learning: A review. *Journal of Electrocardiology* 2019; 57:S70-S74.
- [**Pastore 2019**] Pastore CA, Samesima N, Pereira Filho HG, Tobias N, Madaloso BA, Facin ME. Applicability of the Electro-Vectorcardiogram in Current Clinical Practice. *Arquivos Brasileiros de Cardiologia* 2019; 113:87-99.
- [**Patidar 2017**] Patidar S, Sharma A, Garg N. Automated Detection of Atrial Fibrillation using Fourier-Bessel expansion and Teager Energy Operator from Electrocardiogram Signals. *Computing in Cardiology* 2017; 44.
- [**Pérez 2020**] Pérez Alday EA, Gu A, Shah A, Robichaux C, Wong AKI, Liu C, et al. Classification of 12-lead ECGs: the PhysioNet/Computing in Cardiology Challenge 2020. *Physiological Measurement* 2020; 41.
- [**Petrénas 2012**] Petrénas A, Marozas V, Sörnmo L, Lukosevicius A. An echo state neural network for QRST cancellation during atrial fibrillation. *IEEE Transactions on Biomedical Engineering* 2012; 59:2950-57.

- [**Petrénas 2015a**] Petrénas A, Marozas V, Sörnmo L. Low-complexity detection of atrial fibrillation in continuous long-term monitoring. *Computers in Biology and Medicine* 2015; 65:184-91.
- [**Petrénas 2015b**] Petrénas A, Sörnmo L, Lukoševicius A, Marozas V. Detection of occult paroxysmal atrial fibrillation. *Medical & Biological Engineering & Computing* 2015; 53:287-97.
- [**Philips 2015**] Philips B, Cheng A. Update on the diagnosis and management of arrhythmogenic right ventricular cardiomyopathy. *Current Opinion in Cardiology* 2015; 31:1.
- [**Pilichou 2016**] Pilichou K, Thiene G, Bauce B, Rigato I, Lazzarini E, Migliore F, Marra MP, Rizzo S, Zorzi Z, Daliento L, Corrado D, Basso C. Arrhythmogenic cardiomyopathy. *Orphanet Journal of Rare Diseases* 2016; 11:1-17.
- [**Pujadas 2022**] Pujadas ER, Raisi-Estabragh Z, Szabo L, et al. Atrial fibrillation prediction by combining ECG markers and CMR radiomics. *Scientific Reports* 2022; 12:18876.
- [**Pürerfellner 2014**] Pürerfellner H, Pokushalov E, Sarkar S, Koehler J, Zhou R, Urban L, Hindricks G. P-wave evidence as a method for improving algorithm to detect atrial fibrillation in insertable cardiac monitors. *Heart Rhythm* 2014; 11:1575-83.
- [**Puszkarski 2022**] Puszkarski B, Hryniow K, Sarwas G. Comparison of neural basis expansion analysis for interpretable time series (N-BEATS) and recurrent neural networks for heart dysfunction classification. *Physiological Measurement* 2022; 43:064006.
- [**Quarta 2011**] Quarta G, Muir A, Pantazis A, Syrris P, Gehmlich K, Garcia-Pavia P, Ward D, Sen-Chowdhry S, Elliott PM, McKenna WJ. Familial evaluation in arrhythmogenic right ventricular cardiomyopathy: impact of genetics and revised task force criteria. *Circulation* 2011; 123:2701-9.
- [**Quarta 2012**] Quarta G, Elliot PM. Criteria for Arrhythmogenic Right Ventricular Cardiomyopathy. *Revista Española de Cardiología (English Version)* 2012; 65: 599-605.
- [**Rabiner 1989**] Rabiner L. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE* 1989; 77:257-286.
- [**Rahimeh 2021**] Rahimeh R, Marianne C, Julien O, Fabien L. An Interpretable Hand-Crafted Feature-Based Model for Atrial Fibrillation Detection. *Frontiers in Physiology* 2021; 12.

- [**Raine 2004**] Raine D, Langley P, Murray A, Dunuwille A, Bourke J. Surface atrial frequency analysis in patients with atrial fibrillation. *Journal of Cardiovascular Electrophysiology* 2004; 15:1021-6.
- [**Reyna 2021**] Reyna MA, Sadr N, Perez Alday EA, Gu A, Shah A, Robichaux C, et al. Will Two Do? Varying Dimensions in Electrocardiography: the PhysioNet/Computing in Cardiology Challenge 2021. *Computing in Cardiology* 2021; 48:1-4.
- [**Reyna 2022**] Reyna MA, Sadr N, Perez Alday EA, Gu A, Shah A, Robichaux C, Rad AB, Elola A, Seyedi S, Ansari S, Ghanbari H, Li Q, Sharma A, Clifford GD. Issues in the automated classification of multilead ecgs using heterogeneous labels and populations. *Physiological Measurement* 2022; 43:084001.
- [**Rieta 2004**] Rieta J, Castells F, Sánchez C, Zarzoso V, Millet J. Atrial activity extraction for atrial fibrillation analysis using blind source separation. *IEEE Transactions on Biomedical Engineering* 2004; 51:1176-86.
- [**Rizwan 2021**] Rizwan A, Zoha A, Mabrouk IB, Sabbour HM, Al-Sumaiti AS, Alomainy A, Imran MA, Abbasi QH. A Review on the State of the Art in Atrial Fibrillation Detection Enabled by Machine Learning. *IEEE Reviews in Biomedical Engineering* 2021; 14:219-239.
- [**Ródenas 2015**] Ródenas J, García M, Alcaraz R, Rieta J. Wavelet entropy automatically detects episodes of atrial fibrillation from single-lead electrocardiograms. *Entropy* 2015; 17:6179-99.
- [**Rohr 2022**] Rohr M, Reich C, Höhl A, Lilienthal T, Dege T, Plesinger F, Bulkova V, Clifford G, Reyna M, Hoog-Antink C. Exploring novel algorithms for atrial fibrillation detection by driving graduate level education in medical machine learning. *Physiological Measurement* 2022;43:074001.
- [**Rosenblatt 1958**] Rosenblatt F. The Perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review* 1958; 65:386-408.
- [**Ruan 2011**] Ruan X, Liu C, Wang X, Li P. Automatic Detection of Atrial Fibrillation Using R-R Interval Signal. *4th International Conference on Biomedical Engineering and Informatics (BMEI)* 2011; 644-647.
- [**Ruwald 2015**] Ruwald AC, Marcus F, Estes 3rd NA, Link M, McNitt S, Polonsky B, et al. Association of competitive and recreational sport participation with cardiac events in patients with arrhythmogenic right ventricular cardiomyopathy: results from the North American multidisciplinary study of arrhythmogenic right ventricular cardiomyopath. *European Hear Journal* 2015; 36:1735-43.

- [Sadr 2018] Sadr N, Jayawardhana M, Pham TT, Tang R, Balaei AT, Chazal P. A low-complexity algorithm for detection of atrial fibrillation using an ECG. *Physiological Measurement* 2018; 39:6.
- [Sagie 1992] Sagie A, Larson MG, Goldberg RJ, Bengtson JR, Levy D. An improved method for adjusting the QT interval for heart rate (the Framingham Heart Study). *American Journal of Cardiology* 1992; 70:797-801.
- [Sakoe 1978] Sakoe H, Chiba S. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustic, Speech and Signal Processing* 1978; 26:43-49.
- [Saladin 2012] Saladin K. Anatomía y Fisiología. La unidad entre forma y función, 6a edición. *McGraw-Hill* 2012.
- [Sandberg 2008] Sandberg F, Stridh M, Sornmo L. Frequency Tracking of Atrial Fibrillation Using Hidden Markov Models. *IEEE Transactions on Biomedical Engineering* 2008; 55:502-11.
- [Sanz 2018] Estudio de la Miocardiopatía Arritmogénica en nuestro entorno. Mejora en el diagnóstico de las formas con afectación de ventrículo izquierdo. Tesis doctoral. Universitat de València, 2018.
- [Sarkar 2008] Sarkar S, Ritscher D, Mehra R. A detector for a chronic implantable atrial tachyarrhythmia monitor. *IEEE Transactions on Biomedical Engineering* 2008; 55:1219-24.
- [Sasaki 2007] Sasaki Y. The truth of the F-measure. *University of Manchester* 2007.
- [Savarese 2022] Savarese G, Becher PM, Lund LH, Seferovic P, Rosano G, Coats A. Global burden of heart failure: a comprehensive and updated review of epidemiology. *Cardiovascular Research* 2022; 118:3272–3287.
- [Savelieva 2008] Savelieva I, Camm J. Update on atrial fibrillation: part I. *Clinical Cardiology* 2008; 31:55-62.
- [Sawant 2022] Sawant NK, Patidar S. Application of Fourier-Bessel expansion and LSTM on multi-lead ECG for cardiac abnormalities identification. *Physiological Measurement* 2022; 43:124002.
- [Segura 2020] Segura-Rodríguez D, Bermúdez-Jiménez FJ, Carriel V, López-Fernández S, González-Molina M, Oyonarte Ramírez JM, Fernández-Navarro L, García-Roa MD, Cabrerizo EM, Durand-Herrera D, Alaminos M, Campos A, Macías R, Álvarez M, Tercedor L, Jiménez-Jáimez J. Myocardial fibrosis in

- arrhythmogenic cardiomyopathy: a genotype-phenotype correlation study. *European Heart Journal Cardiovascular Imaging* 2020; 21:378–386.
- [**Sen 2008**] Sen-Chowdhry S et al. Left-dominant arrhythmogenic cardiomyopathy: an under-recognized clinical entity. *Journal of the American College of Cardiology* 2008; 2175-87.
- [**Shaffer 2017**] Shaffer F, Ginsberg J. An Overview of Heart Rate Variability Metrics and Norms. *Frontiers in Public Health* 2017; 5:1-17.
- [**Sharma 2022**] Sharma A, Bosman LP, Tichnell C, Nanavati J, Murray B, Nonyane BAS, Tandri H, Calkins H, James CA. Arrhythmogenic Right Ventricular Cardiomyopathy Prevalence and Arrhythmic Outcomes in At-Risk Family Members: A Systematic Review and Meta-Analysis. *Circulation: Genomic and Precision Medicine* 2022; 15:e003530.
- [**Simov 2016**] Simov D. Electrocardiographic changes in certain cardiovascular physiological and pathological settings. Impact on coronary artery bypass grafting. *International Journal Bioautomation* 2016; 20:43–68.
- [**Slocum 1992**] Slocum J, Sahakian A, Swiryn S. Diagnosis of atrial fibrillation from surface electrocardiograms based on computer-detected atrial activity. *Journal of Electrocardiology* 1992; 25:1-8.
- [**Sodhro 2019**] Sodhro AH, Luo Z, Sodhro GH, Muzamal M, Rodrigues JJ, Albuquerque VH. Artificial intelligence based QOS optimization for multimedia communication in iov systems. *Future Generation Computer Systems* 2019; 95:667-680.
- [**Sodmann 2018**] Sodmann P, Vollmer M, NathNand Kaderali L. A convolutional neural network for ECG annotation as the basis for classification of cardiac rhythms *Physiological Measurement* 2018; 39 104005.
- [**Spaccarotella 2020**] Spaccarotella CAM, Polimeni A, Migliarino S, Principe E, Curcio A, Mongiardo A, Sorrentino S, De Rosa S, Indolfi C. Multichannel Electrocardiograms Obtained by a Smartwatch for the Diagnosis of ST-Segment Changes. *JAMA Cardiology* 2020; 5:1176-1180.
- [**Stridh 2001**] Stridh M, Sornmo L, Meurling C, Olsson S. Characterization of atrial fibrillation using the surface ECG: time-dependent spectral properties. *IEEE Transactions on Biomedical Engineering* 2001; 48:19-27.
- [**Suárez 2013**] Suárez-Mier MP et al. Pathology of sudden death during recreational sports in Spain. *Forensic Science International* 2013; 10:188-96.

- [Sweet 2015] Sweet M, Taylor MR, Mestroni L. Diagnosis, prevalence, and screening of familial dilated cardiomyopathy. *Expert Opin Orphan Drugs* 2015; 3:869-876.
- [Tadi 2019] Tadi MJ, et al. Comprehensive analysis of cardiogenic vibrations for automated detection of atrial fibrillation using smartphone mechanocardiograms. *IEEE Sensors Journal* 2019; 19:2230-2242.
- [Teijeiro 2017] Teijeiro T, García CA, Castro D, Félix P. Arrhythmia Classification from the Abductive Interpretation of Short Single-Lead ECG Records. *Computing in Cardiology* 2017; 44.
- [Tihonenko 2008] Tihonenko V, Khaustov A, Ivanov S, Rivin A, Yakushenko E. St Petersburg INCART 12-lead arrhythmia database. *PhysioBank PhysioToolkit PhysioNet* 2008; doi:10.13026/C2V88N.
- [Titomir 2007] Titomir LI, Trunov VG, Aidu E, Sakhnova TA, Blinova EV. New approaches to the diagnosis of left and right ventricular hypertrophy by means of dipolar electrocardiotopography. *Anatolian Journal of Cardiology/Anadolu Kardiyoloji Dergisi*, 2007; 7.
- [Tomé 2004] Tomé MT, García-pinilla JM, McKenna J. Actualización en miocardiopatía arritmogénica del ventrículo derecho: genética, diagnóstico, manifestaciones clínicas y estratificación de riesgo. *Revista Española de Cardiología* 2004; 57:757–67.
- [Tortajada 2012] Tortajada S, García-Gómez J, Robles-Viejo M. Incremental Learning approaches to Biomedical decision problems. *Universitat Politècnica de València* 2012.
- [Touahria 2021] Touahria R, Hacine-Gharbi A, Ravier P. Discrete Wavelet based Features for PCG Signal Classification using Hidden Markov Models. *10th International Conference on Pattern Recognition Applications and Methods (ICPRAM)* 2021; 334-340.
- [Towbin 2019] Towbin JA, McKenna WJ, Abrams DJ, Ackerman MJ, Calkins H, Darrieux FCC, Daubert JP, Chillou C, et al. HRS Expert consensus statement on evaluation, risk stratification, and management of arrhythmogenic cardiomyopathy. *Heart Rhythm* 2019; 16:e373–e407.
- [Tsuji 1996] Tsuji H, Larson M, Venditti F, Manders E, Evans J, Feldman C, Levy D. Impact of Reduced Heart Rate Variability on Risk for Cardiac Events: The Framingham Heart Study. *Circulation* 1996; 94:2850-55.

- [**Vaduganathan 2022**] Vaduganathan M, Mensah GA, Varieur Turco J, Fuster V, Roth GA. The Global Burden of Cardiovascular Diseases and Risk: A Compass for Future Health. *Journal of the American College of Cardiology* 2022; 80:2361–2371.
- [**Vafaie 2014**] Vafaie M, Ataei M, Koofigar H. Heart diseases prediction based on ECG signals' classification using a genetic-fuzzy system and dynamical model of ECG signals. *Biomedical Signal Processing and Control* 2014; 14:291–6.
- [**Vásquez 2001**] Vásquez C, Hernandez A, Mora F, Carrault G, Passariello G. Atrial activity enhancement by wiener filtering using an artificial neural network. *IEEE Transactions on Biomedical Engineering* 2001; 48:940-4.
- [**Vázquez 2022**] Vázquez CG, Breuss A, Gnarra O, Portmann J, Madaffari A, Da Poian G. Label noise and self-learning label correction in cardiac abnormalities classification. *Physiological Measurement* 2022; 43:094001.
- [**Virani 2021**] Virani SS et al. Heart disease and stroke statistics—2021 Update: a report from the American Heart Association. *Circulation* 2021; 143:e254.
- [**Wagner 2020**] Wagner P, Strodthoff N, Boussejot RD, Kreiseler D, Lunze F I, Samek W, Schaeffter T. PTB-XL, a large publicly available electrocardiography dataset. *Scientific Data* 2020; 7:1–15.
- [**Wang 2004**] Wang TJ, Parise H, Levy D, D'Agostino RB, Wolf PA, Vasani RS, Benjamin EJ. Obesity and the risk of new-onset atrial fibrillation. *JAMA* 2004; 292:2471-7.
- [**Wang 2015**] Wang S, Petzold M, Cao J, Zhang Y, Wang W. Direct Medical Costs of Hospitalizations for Cardiovascular Diseases in Shanghai, China. Trends and Projections. *Medicine* 2015; 94e837.
- [**Wheeldon 1995**] Wheeldon N. Atrial fibrillation and anticoagulant therapy. *European Heart Journal*, 1995; 16:302-12.
- [**Wickramasinghe 2021**] Wickramasinghe NL, Athif M. Multi-label Cardiac Abnormality Classification from Electrocardiogram using Deep Convolutional Neural Networks. *Computing in Cardiology* 2021; 48.
- [**Winters 2007**] Winters-Hilt S, Merat S. SVM clustering. *BMC Bioinformatics* 2007; 8:S18.
- [**Xia 2022**] Xia P, He Z, Bai Z, Wang Y, Yu X, Geng F, Du L, Chen X, Wang P, Zhu Y. A novel multi-scale 2D CNN with weighted focal loss for arrhythmias

- detection on varying-dimensional ECGs. *Physiological Measurement* 2022; 43:104003.
- [**Xiaoyu 2021**] Xiaoyu L, Chen L, Xian X, Yuhua W, Jishang W, Yuyao S, Buyue Q, Xiao X. Towards Generalization of Cardiac Abnormality Classification Using ECG Signal. *Computing in Cardiology* 2021; 48.
- [**Xiong 2017**] Xiong Z, Stiles MK, Jichao Zhao J. Robust ECG Signal Classification for the Detection of Atrial Fibrillation Using Novel Neural Networks. *Computing in Cardiology* 2017; 44.
- [**Xiong 2018**] Xiong Z, NashMP, Cheng E, Fedorov VV, Stiles MK and Zhao J. Ecg signal classification for the detection of cardiac arrhythmias using a convolutional recurrent neural network. *Physiological. Measurement* 2018; 39 094006.
- [**Xu 2022**] Xu Z, Guo Y, Zhao T, Zhao Y, Liu Z, Sun X, Xie G, Li Y. Abnormality classification from electrocardiograms with various lead combinations. *Physiological Measurement* 2022; 43:074002.
- [**Yang 2013**] Yang S, Shen H. Heartbeat Classification using discrete wavelet transform and kernel principal component analysis. *IEEE 2013 Tencon - Spring* 2013; 34-38.
- [**Yang 2021**] Yang X, Zhang X, Yang M, Zhang L. 12-Lead ECG arrhythmia classification using cascaded convolutional neural network and expert feature. *Journal of Electrocardiology* 2021; 67:56-62.
- [**Yin 2019**] Yin L, Chen F, Zhang Q, Ma X. Arrhythmia classification based on multi-domain feature extraction. *Journal of Physics: Conference Series* 2019; 1237:022062.
- [**Young 2009**] Young S, et al. The HTK Book. *Cambridge University Press* 2009.
- [**Zabihi 2017**] Zabihi M, Ali Bahrami A, Katsaggelos AK, Kiranyaz S, Narkilahti S, Gabbouj M. Detection of Atrial Fibrillation in ECG Hand-held Devices Using a Random Forest Classifier. *Computing in Cardiology* 2017; 44.
- [**Zeiler 2013**] Zeiler M, Ranzato M, Monga R. On rectified linear units for speech processing. *Speech and Signals Processing (ICASSP)* 2013; 1:3517-21.
- [**Zhang 2021**] Zhang D, Yang Y, Yuan X, Zhang P. Interpretable deep learning for automatic diagnosis of 12-lead electrocardiogram. *iScience*, 2021; 24:102373.

- [**Zhang 2015**] Zhang L, Guo T, Xi B, Fan Y, Wang Y. Automatic recognition of cardiac arrhythmias based on the geometric patterns of Poincaré plots. *Physiological Measurement* 2015; 36:283-301.
- [**Zhang 2023**] Zhang H, Liu C, Tang F, Liu F, et al. Atrial fibrillation classification based on the 2D representation of minimal subset ECG and a non-deep neural network. *Frontiers in Physiology* 2023; 14:1-11.
- [**Zhao 2022**] Zhao Z et al. Analysis of an adaptive lead weighted ResNet for multiclass classification of 12-lead ECGs. *Physiological Measurement*, 2022; 43 034001.
- [**Zheng 2020a**] Zheng J, Chu H, Struppa D et al. Optimal Multi-Stage Arrhythmia Classification Approach. *Scientific Reports* 2020; 10 2898.
- [**Zheng 2020b**] Zheng J, Cui H, Struppa D, Zhang J, Yacoub SM, El-Askary H, et al. Optimal Multi-Stage Arrhythmia Classification Approach. *Scientific Data* 2020; 10:1–17.
- [**Zheng 2020c**] Zheng J, Zhang J, Danioko S, Yao H, Guo H, Rakovski C. A 12-lead Electrocardiogram Database for Arrhythmia Research Covering More Than 10,000 Patients. *Scientific Data* 2020; 7:1–8.
- [**Zhou 2014**] Zhou X, Ding H, Ung B, Pickwell-MacPherson E, Zhang Y. Automatic online detection of atrial fibrillation based on symbolic dynamics and Shannon entropy. *BioMedical Engineering OnLine* 2014; 13.
- [**Zhu 2021**] Zhu Z *et al.* Identification of 27 abnormalities from multi-lead ECG signals: an ensembled SE_ResNet framework with Sign Loss function. *Physiological Measurement* 2021; 42 065008.
- [**Zihlmann 2017**] Zihlmann M, Perekrestenko D, Tschannen M. Convolutional Recurrent Neural Networks for Electrocardiogram Classification. *Computing in Cardiology* 2017; 44.