



UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA

*DSIC*

DEPARTAMENTO DE SISTEMAS  
INFORMÁTICOS Y COMPUTACIÓN



**MIARFID**

Máster en Inteligencia Artificial,  
Reconocimiento de Formas  
e Imagen Digital

# **Clasificación automática de registros ECG para la detección de Fibrilación Auricular y otros ritmos cardiacos**

**TRABAJO FIN DE MÁSTER**  
**Trabajo de investigación**

*Autor*

Santiago Jiménez Serrano

*Directores*

Carlos D. Martínez Hinarejos  
Francisco Sales Castells Ramón

*València, Julio de 2018*



# Resumen

---

La importancia clínica de las arritmias cardiacas está aumentando, junto con su incidencia y prevalencia, principalmente asociadas con el envejecimiento de la población. Entre estas enfermedades destaca la Fibrilación Auricular (FA) ya que es el tipo de arritmia sostenida más común en adultos con una tendencia creciente más significativa, siendo en muchas ocasiones difícil de diagnosticar debido a un comportamiento paroxístico y/o la ausencia de síntomas en algunos casos.

Por otro lado, hoy en día estamos en un escenario en el que los dispositivos portátiles o “*wearables*” están ganando gran interés como dispositivos de monitorización, tanto en investigación como en ámbitos clínicos. Sin embargo, los métodos automáticos para proporcionar un diagnóstico fiable de la FA utilizando las señales de electrocardiograma (ECG) proporcionadas por dispositivos portátiles son todavía un reto, especialmente si también se consideran otros ritmos normales o patológicos.

El objetivo de este Trabajo Final de Máster es proporcionar diversos modelos de clasificación junto con su rendimiento para discriminar registros cortos de ECG de una única derivación entre cuatro grupos: ritmo normal (N), FA (A), otros ritmos (O) y ruidoso (~).

Como base de datos para este estudio se utilizaron 8.528 registros de ECG de una única derivación con duraciones entre 9 y 60 segundos, proporcionados en el contexto de la competición *2017 PhysioNet/Computing in Cardiology Challenge*.

La estrategia propuesta en este trabajo se basa inicialmente en la extracción automática de características derivadas de la actividad ventricular de las señales ECG. Posteriormente se realizó una selección de características utilizando dos metodologías distintas: *Backward Elimination* y *Forward Selection*. Finalmente, las características seleccionadas se utilizaron para entrenar y evaluar mediante validación cruzada el rendimiento de diferentes modelos de clasificación, principalmente redes neuronales de tipo *feedforward* (FFNN), así como modelos Naïve Bayes y *Support Vector Machine* (SVM). A cada uno de estos modelos se le realizó un ajuste de parámetros de entrenamiento mediante *grid-search* durante la fase de validación.

Los resultados mostraron que los modelos que presentaban mejor rendimiento de clasificación fueron las redes neuronales ( $F_1 = 0.75$ ), seguidas de cerca por los modelos SVM ( $F_1 = 0.73$ ), mientras que Naïve Bayes presentó el menor rendimiento ( $F_1 = 0.67$ ). Asimismo, también se demostró que la mayor dificultad de este problema se encuentra en la identificación de otros ritmos anómalos distintos a la fibrilación auricular, así como de los registros ruidosos.

Dado que las señales utilizadas comparten muchas características con las adquiridas con dispositivos móviles, los modelos de clasificación resultantes podrían ser buenos candidatos para ser implementados en sistemas de gestión de pacientes con dispositivos *wearables*, ya que este enfoque tiene un bajo consumo computacional durante la clasificación.

---

## Palabras Clave

Redes Neuronales, Naïve Bayes, *Support Vector Machine*, extracción de características, selección de características, clasificación, ECG, Fibrilación Auricular, intervalo RR.

# Resum

---

La importància clínica de les arítmies cardíques està augmentant, juntament amb la seua incidència i prevalença, principalment associades amb l'envelliment de la població. Entre aquestes malalties destaca la Fibril·lació Auricular (FA) ja que és el tipus d'arítmia sostinguda més comuna en adults amb una tendència creixent més significativa, essent en moltes ocasions difícil de diagnosticar a causa d'un comportament paroxíctic i/o l'absència de símptomes en alguns casos.

D'altra banda, avui dia ens trobem a un escenari en el qual els dispositius portàtils o "wearables" estan guanyant gran interès com a dispositius de monitoratge, tant en recerca com en àmbits clínics. No obstant això, els mètodes automàtics per a proporcionar un diagnòstic fiable de la FA utilitzant els senyals d'electrocardiograma (ECG) proporcionades pels dispositius portàtils són encara un repte, especialment si també es consideren altres ritmes normals o patològics.

L'objectiu d'aquest Treball Final de Màster és proporcionar diversos models de classificació juntament amb el seu rendiment per a discriminar registres curts d'ECG d'una única derivació entre quatre grups: ritme normal (N), FA (A), altres ritmes (O) i sorolls (~).

Com a base de dades per a aquest estudi es van utilitzar 8.528 registres d'ECG d'una única derivació amb durades entre 9 i 60 segons, proporcionats en el context de la competició *2017 PhysioNet/Computing in Cardiology Challenge*.

L'estratègia proposada a aquest treball es basa inicialment en l'extracció automàtica de característiques derivades de l'activitat ventricular dels senyals ECG. Posteriorment es va realitzar una selecció de característiques utilitzant dues metodologies diferents: *Backward Elimination* i *Forward Selection*. Finalment, les característiques seleccionades es van utilitzar per a entrenar i avaluar mitjançant validació creuada el rendiment de diferents models de classificació, principalment xarxes neuronals de tipus *feedforward* (FFNN), així com models Naïve Bayes i *Support Vector Machine* (SVM). A cadascun d'aquests models se li va realitzar un ajust de paràmetres d'entrenament *mitjançant grid-search* durant la fase de validació.

Els resultats van mostrar que els models que presentaven millor rendiment de classificació van ser les xarxes neuronals ( $F_1 = 0.75$ ), seguides d'a prop pels models SVM ( $F_1 = 0.73$ ), mentre que Naïve Bayes va presentar el menor rendiment ( $F_1 = 0.67$ ). Així mateix, també es va demostrar que la major dificultat d'aquest problema es troba en la identificació d'altres ritmes anòmals diferents a la fibril·lació auricular, així com dels registres sorollosos.

Atès que els senyals utilitzats comparteixen moltes característiques amb les adquirides amb dispositius mòbils, els models de classificació resultants podrien ser bons candidats per a ser implementats en sistemes de gestió de pacients amb dispositius *wearables*, pel fet que aquesta aproximació té un baix consum computacional durant la classificació.

---

## Paraules Clau

Xarxes Neuronals, Naïve Bayes, *Support Vector Machine*, extracció de característiques, selecció de característiques, classificació, ECG, Fibril·lació Auricular, interval RR.

# Abstract

---

The clinical importance of cardiac arrhythmias is increasing, along with its incidence and prevalence, mainly associated with the aging of the population. Among these diseases Atrial Fibrillation (AF) stands out since it is the type of sustained arrhythmia most common in adults with a more significant growing tendency, being in many cases difficult to diagnose due to a paroxysmal behavior and/or the absence of symptoms in some patients.

On the other hand, today we are in a scenario in which mobile devices or “wearables” are gaining great interest as monitoring devices, both in research and in clinical settings. However, automatic methods to provide a reliable diagnosis of AF using electrocardiogram signals (ECG) provided by mobile devices are still a challenge, especially if other normal or pathological rhythms are also considered.

The main objective of this Final Master's Thesis is to provide different classification models together with their performance to discriminate short ECG single-lead records among four different groups: normal rhythm (N), FA (A), other rhythms (O) and noisy (~).

As database for this study, 8,528 single-lead ECG records lasting among 9 and 60 seconds were used, provided in the context of the *2017 PhysioNet/Computing in Cardiology Challenge*.

The proposed strategy in this work is initially based on the automatic extraction of features mainly derived from the ventricular activity of the ECG signals. Next, a selection of characteristics was made using two different methodologies: *Backward Elimination* and *Forward Selection*. Finally, the selected features were used to train and evaluate through cross-validation the performance of different classification models, mainly feedforward neural networks (FFNN), as well as Naïve Bayes and Support Vector Machine (SVM) models. The training parameters for each of these models were tuned through a grid-search validation process.

Results showed that the models with the best classification performance were the neural networks ( $F_1 = 0.75$ ), followed closely by the SVM models ( $F_1 = 0.73$ ), while Naïve Bayes presented the lowest performance ( $F_1 = 0.67$ ). Likewise, it was also proved that the greatest difficulty of this problem lies on the identification of other anomalous rhythms other than atrial fibrillation, as well as in the noisy registers.

Since the signals used share many characteristics with those acquired with mobile devices, the resulting classification models could be good candidates to be implemented in patient management systems with wearable devices, since this approach has a low computational consumption during classification.

---

## Keywords

Neural Networks, Naïve Bayes, Support Vector Machine, feature extraction, feature selection, classification, ECG, Atrial Fibrillation, RR interval.



## Abreviaturas

---

AA	Aprendizaje Automático
ANN	<i>Artificial Neural Networks</i> - Redes Neuronales Artificiales
AVN	Nodo Aurículoventricular
ECG	Electrocardiograma
FA	Fibrilación Auricular
FFNN	Redes neuronales <i>feedforward</i>
HMM	<i>Hidden Markov Models</i> – Modelos ocultos de Markov
IA	Inteligencia Artificial
KDF	<i>Kernel Density Function</i> - Función de densidad kernel
KNN	<i>K-Nearest Neighbors</i> – K-Vecinos más cercanos
NB	Naïve Bayes
PA	Potencial de Acción
PCA	Análisis de Componentes Principales
RBF	Función de base radial
RNA	Redes Neuronales Artificiales
RP	Reconocimiento de Patrones
SADM	Sistemas de Ayuda a la Decisión Médica
SAN	Nodo Sinusal
SVM	<i>Support Vector Machines</i>
VC	Validación Cruzada

# Índice de Contenidos

---

1. Introducción .....	1
1.1. Objetivos .....	2
2. Antecedentes clínicos .....	3
2.1. Fibrilación Auricular .....	6
2.2. Otras arritmias cardíacas.....	8
3. Estado del arte.....	9
3.1. Detección de FA basada en la actividad auricular .....	9
3.2. Detección de FA basada en la actividad ventricular .....	11
3.3. <i>Body Surface Potential Mapping</i> .....	13
3.4. Simulaciones computacionales .....	14
4. Marco teórico .....	17
4.1. Aprendizaje Automático y Reconocimiento de Patrones.....	17
4.2. Redes Neuronales Artificiales.....	21
4.3. Naïve Bayes .....	24
4.4. <i>Support Vector Machines</i> .....	26
5. Materiales .....	29
5.1. Registros ECG .....	29
5.2. <i>Gold Standard</i> de los registros ECG.....	32
6. Métodos.....	35
6.1. Criterios de evaluación .....	35
6.2. Preprocesado de señal.....	37
6.3. Extracción de características.....	38
6.4. Preprocesado de las características del conjunto de datos .....	43
6.5. Selección de características .....	44
6.6. Entrenamiento y validación de las Redes Neuronales .....	45
6.7. Entrenamiento y validación de Naïve Bayes.....	46
6.8. Entrenamiento y validación de SVM.....	47
7. Resultados.....	49
7.1. Selección de características .....	49
7.2. Redes neuronales .....	51
7.3. Naïve Bayes .....	54
7.4. SVM .....	56
7.5. Visión global de los resultados .....	59
8. Discusión .....	61

9. Conclusiones .....	65
10. Trabajos futuros.....	67
11. Bibliografía.....	69
Apéndice A – Publicación .....	78
Apéndice B – Ranking <i>Physionet Computing in Cardiology Challenge 2017</i> .....	84



# 1. Introducción

---

Las enfermedades cardiovasculares son una de las principales causas de mortalidad y morbilidad en el mundo. Estas enfermedades disminuyen la calidad de vida del paciente y ocasionan un coste elevado a los sistemas Nacionales de Salud [1]. En este sentido, la importancia clínica de las arritmias cardiacas está aumentando, junto con su incidencia y prevalencia, al estar principalmente asociadas con el envejecimiento de la población [2]. Entre estas enfermedades destaca la Fibrilación Auricular (FA) [3], ya que es el tipo de arritmia sostenida más común en adultos, además de poseer la tendencia creciente más significativa, sobre todo en población de edad avanzada o con trastornos de obesidad [4]. Además, en ocasiones la FA es difícil de diagnosticar debido a comportamientos paroxísticos y/o la posible ausencia de síntomas [5].

Por otro lado, hoy en día los dispositivos portátiles (también conocidos como *wearables*) están ganando gran interés como dispositivos de monitorización tanto en investigación como en ámbitos clínicos [6]. Sin embargo, los métodos automáticos para proporcionar un diagnóstico fiable de la FA utilizando las señales de electrocardiograma (ECG) obtenidas con dispositivos portátiles son todavía un reto, especialmente si también se consideran otros ritmos normales o patológicos.

En este sentido, los Sistemas de Ayuda a la Decisión Médica (SADM) proveen conocimiento específico y preciso para la toma de decisiones a adoptar durante pronósticos, diagnósticos, tratamientos y gestión de pacientes. Los SADM están relacionados con el concepto de medicina basada en la evidencia [7], ya que infieren conocimiento a partir de bases de datos clínicas, para posteriormente ayudar en el diagnóstico de nuevos pacientes utilizando el conocimiento adquirido.

Los SADM más complejos están basados en el uso de la Inteligencia Artificial (IA) en medicina [8], en concreto con la rama de Aprendizaje Automático (AA) o *Machine Learning* [9], donde se utilizan algoritmos para obtener modelos predictivos a partir de los datos clínicos de entrada.

Por otra parte, el Reconocimiento de Patrones (RP) o *Pattern Recognition* es la siguiente etapa natural del AA, donde los modelos predictivos resultantes asignan una etiqueta o clase a un conjunto de valores de entrada del que inicialmente se desconoce a qué grupo pertenece. En los problemas de clasificación, estas etiquetas o clases son conocidas *a priori*.

En este trabajo, el ejemplo más directo de *Pattern Recognition* sería determinar si una señal ECG corresponde a la de una persona que padece FA, cualquier otro ritmo patológico o, por el contrario, pertenece a un individuo con un ritmo cardiaco normal.

## 1.1. Objetivos

El objetivo principal de este Trabajo Final de Máster es proporcionar diversos modelos de clasificación y evaluar su rendimiento para clasificar registros cortos de ECG de una única derivación entre cuatro grupos diferentes: ritmo normal (N), FA (A), otros ritmos (O) y ruidoso (~). Para ello se utilizará la base de datos de señales ECG de una única derivación proporcionada en el contexto del *Challenge 2017 PhysioNet/Computing in Cardiology* [10].

La estrategia propuesta en este trabajo se basa en la sucesión de las siguientes etapas:

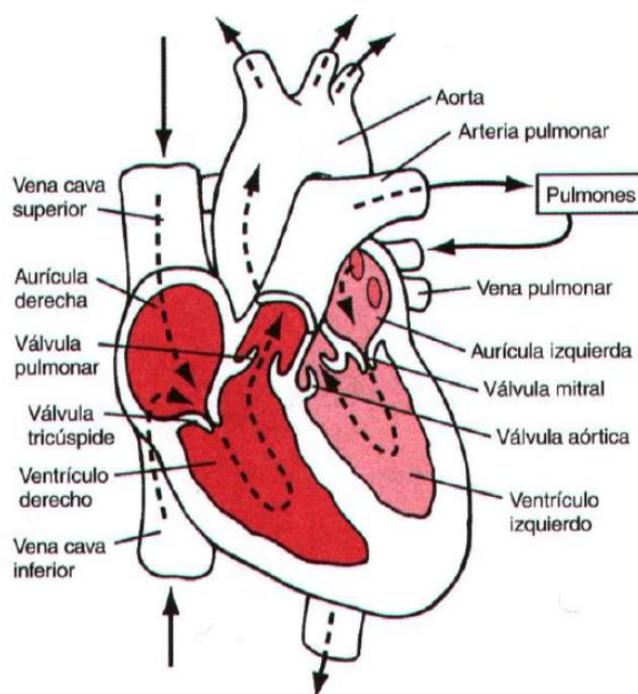
1. Revisión bibliográfica de las características extraídas de la señal ECG y de las técnicas utilizadas en otros trabajos para la detección de la FA.
2. Extracción automática de características derivadas de la actividad ventricular de las señales ECG disponibles.
3. Selección de características utilizando las metodologías *Backward-Elimination* y *Forward-Selection* respectivamente.
4. Finalmente, las características seleccionadas se usarán para entrenar y evaluar mediante validación cruzada (VC) el rendimiento de clasificación de un conjunto de diferentes modelos de clasificación, principalmente redes neuronales de tipo *Feedforward* (FFNN), así como modelos Naïve Bayes (NB) y *Support Vector Machine* (SVM). A cada uno de estos modelos se le realizará un ajuste de parámetros de entrenamiento mediante una *grid-search* durante la fase de validación, seleccionando así aquellos que presenten un mejor rendimiento de clasificación.

Nótese que las señales que utilizaremos durante este trabajo comparten muchas características con las adquiridas por distintos dispositivos *wearables* en el mercado. Es por esto por lo que los modelos de clasificación resultantes podrían ser buenos candidatos para ser implementados en sistemas de gestión de pacientes o SADM que utilicen dispositivos portátiles para la monitorización cardiaca, ya que este enfoque presenta un bajo consumo computacional durante la clasificación.

## 2. Antecedentes clínicos

---

El corazón se engloba en el sistema circulatorio, donde actúa como una bomba electromecánica distribuyendo la sangre que provee oxígeno y nutrientes a cada una de las células del cuerpo. Este órgano se divide principalmente en cuatro partes: superior, inferior, izquierda y derecha respectivamente. Las dos cavidades de la parte superior son llamadas aurículas, mientras que las inferiores corresponden a los ventrículos. La Figura 1 muestra la anatomía de un corazón.



**Figura 1. Anatomía del corazón dividida en 4 cavidades principales: dos superiores llamadas aurículas y dos inferiores llamadas ventrículos. La sangre transportada desde las grandes venas desemboca en las aurículas y de éstas pasa a los ventrículos a través de las válvulas auriculoventriculares. Tras la contracción ventricular, la sangre es expulsada hacia los pulmones y los órganos periféricos respectivamente [11].**

Por otro lado, los cardiomiocitos son las células musculares encargadas de la contracción del tejido cardíaco, y se encuentran tanto en las aurículas como en los ventrículos. Durante los latidos, los ventrículos realizan el mayor esfuerzo mecánico, por lo que es aquí donde se encuentra una concentración de cardiomiocitos mayor que en el resto del corazón, y por lo tanto mayor tejido muscular (ver Figura 2).

Además, existe una red que conduce y sincroniza los impulsos eléctricos a través del haz de His y las fibras de Purkinje desde la zona auricular a la ventricular. Esta propagación de la despolarización y la repolarización eléctrica que tiene lugar durante cada latido del corazón produce cambios de voltaje que se pueden medir usando electrodos colocados en la superficie del cuerpo. Estos cambios eléctricos producen una señal comúnmente conocida como electrocardiograma (ECG). La Figura 3 muestra una representación de los elementos más importantes que intervienen en la generación de un complejo P-QRS-T del ECG, o lo que es lo mismo, durante un latido o ciclo cardíaco.

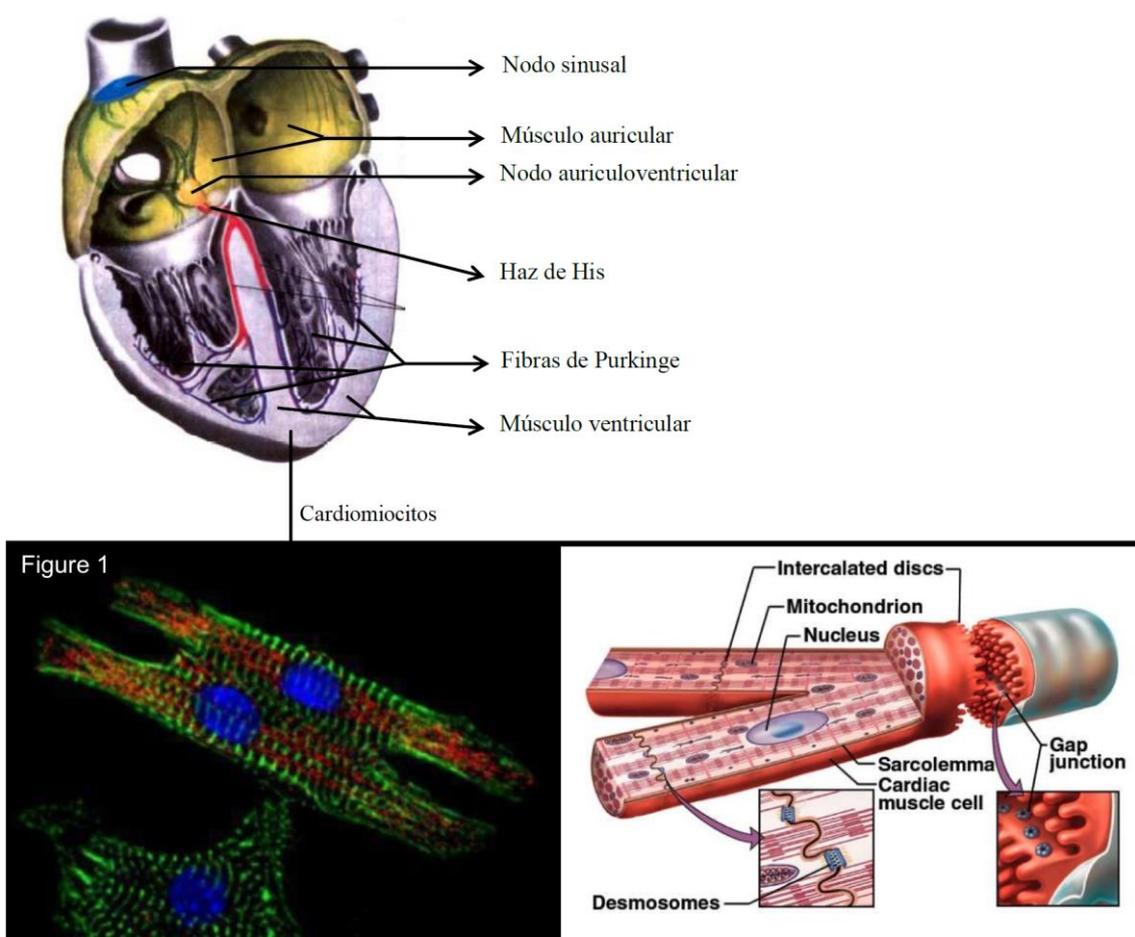


Figura 2. Red de conducción eléctrica del corazón, junto con las células musculares (cardiomiocitos) que responden mecánicamente a los impulsos eléctricos. Abajo aparece una imagen de inmunofluorescencia de cardiomiocitos en cultivo (izquierda) y una representación de la estructura de los cardiomiocitos (derecha) [12].

Cada latido es producido por la contracción acompañada de los cardiomiocitos auriculares y ventriculares. El nodo sinusal (SAN) junto con el nodo auriculoventricular (AVN) actúan como marcapasos naturales del corazón estimulando el sistema de fibras de Purkinje que finalmente dispara los potenciales de acción (PA) producidos en los cardiomiocitos ventriculares [11]. Son precisamente estos potenciales de acción celulares, junto con el ritmo del nodo sinusal, la base de la propagación eléctrica en el corazón. La Figura 4 muestra con más detalle la relación entre un PA en una célula cardiaca ventricular y su correspondiente respuesta mecánica de contracción, sin la que no sería posible los latidos cardiacos.

Estos PA ventriculares dan lugar al complejo más prominente del ECG, denominado QRS, donde R indica el pico más elevado de la señal. Es por esto que el intervalo R-R sirve típicamente para indicar el ritmo al que late el corazón. Por otro lado, la onda P indica la activación auricular, mientras que la onda T indica la repolarización de los ventrículos.

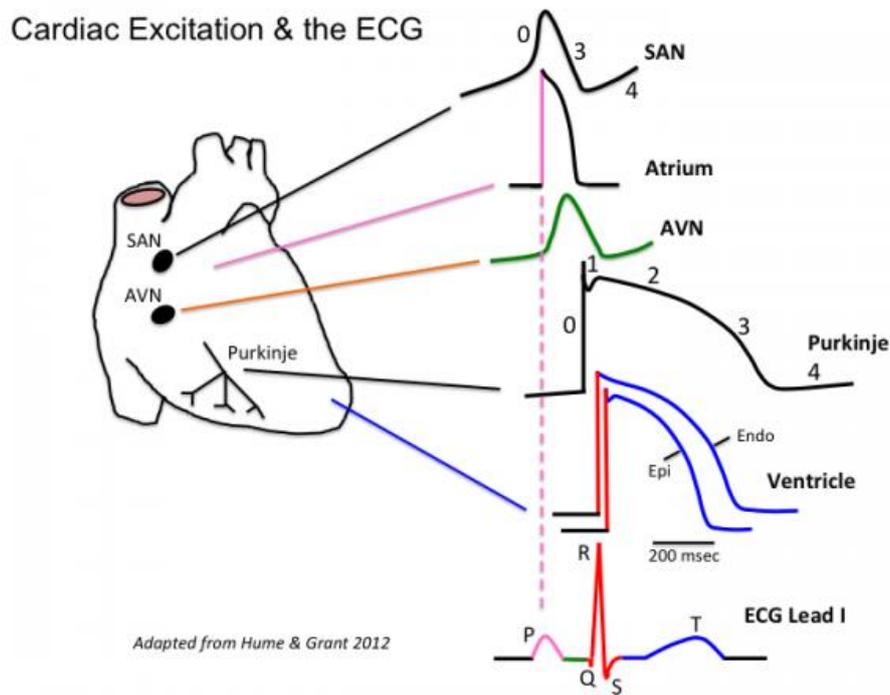


Figura 3. Representación esquemática del corazón, el electrocardiograma (ECG), y los potenciales de acción (PA) registrados en diferentes regiones cardíacas. SAN: nodo sinusal, AVN: nodo auriculoventricular. SAN y AVN actúan como marcapasos naturales. Los diferentes componentes del ECG reflejan diferentes eventos en el ciclo cardíaco, tal como se muestra en la codificación por colores [13].

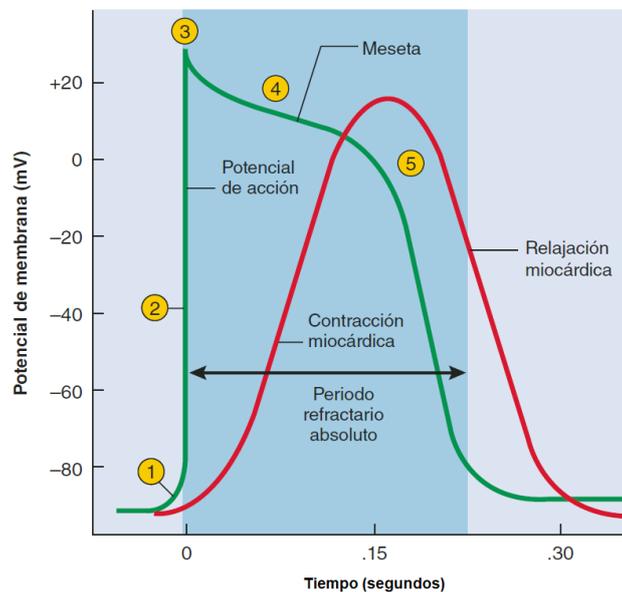


Figura 4. Esquema básico de un potencial de acción (PA) en un cardiomiocito con su correspondiente respuesta mecánica. La curva verde es el PA, mientras que la curva roja representa el aumento y la caída de la tensión muscular mientras el miocardio se contrae y relaja. Durante el período refractario la célula no responde a otros estímulos eléctricos externos [12].

Queda por lo tanto demostrado, con todo lo mencionado anteriormente, la intrínseca relación que presenta un ritmo cardíaco adecuado con la correcta propagación eléctrica en el corazón.

## 2.1. Fibrilación Auricular

La Fibrilación Auricular (FA) es la arritmia cardiaca sostenida más frecuente en adultos, con una prevalencia en torno al 1-2% de la población general [14] [15], y de hasta un 10% en mayores de 75 años [4]. Esta enfermedad está relacionada con un riesgo significativo de mortalidad y morbilidad por asociación de riesgo de muerte, accidente cerebrovascular, hospitalización, insuficiencia cardiaca y enfermedad coronaria, entre otros [16] [17]. Se calcula que más de 12 millones de europeos y norteamericanos sufren de FA, y su prevalencia probablemente se triplicará en los próximos 30-50 años [17]. Más importante aún, la incidencia de FA aumenta con la edad, de menos de 0,5% a los 40-50 años de edad, a 5-15% para los 80 años de edad [5]. Es por ello que la FA se diagnostica cada vez más en la práctica clínica, debido al envejecimiento de la población en países occidentales. En consecuencia, la FA se está convirtiendo en un problema médico cada vez más importante.

Diversas instituciones cardiológicas como el *American College of Cardiology*, la *American Heart Association* y la *European Society of Cardiology* definen la FA como una taquiarritmia caracterizada por una activación auricular predominantemente no coordinada con el consecuente deterioro de la función mecánica auricular [18]. La Figura 5 muestra ejemplos de señales ECG correspondientes a distintas arritmias cardiacas, entre ellas la FA.

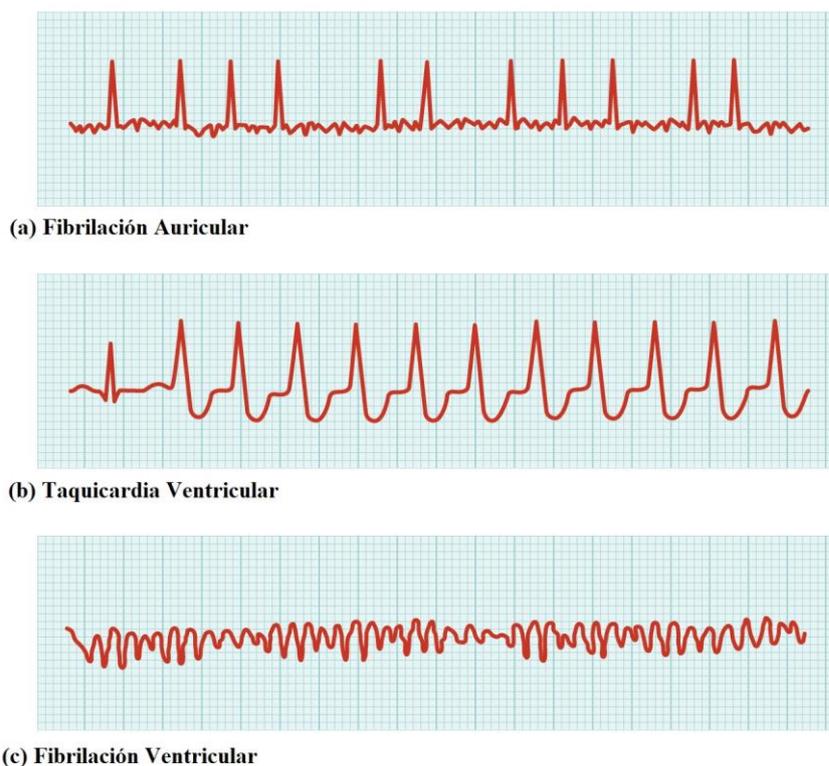


Figura 5. ECGs correspondientes a distintas arritmias cardiacas. En la Fibrilación Auricular (a) existe un ritmo R-R irregular y ausencia de onda P. La Taquicardia Ventricular (b) provoca contracciones muy rápidas produciendo una forma inusual del complejo QRS. La Fibrilación Ventricular (c) es incompatible con la vida, ya que desaparece cualquier coordinación mecánica cardiaca para producir el bombeo de sangre a otros órganos [19].

En contraste con el ritmo sinusal normal, donde un único frente de onda de activación eléctrica se propaga desde las aurículas hacia los ventrículos, la FA se manifiesta por la presencia de múltiples frentes de onda errantes con diferentes patrones de propagación [20]. Dichos frentes de onda aparecen con frecuencias excesivamente altas a un ritmo que predomina sobre el ritmo sinusal, lo que hace que el corazón se contraiga inadecuadamente. En casos extremos, esta frecuencia es tan elevada que los tejidos cardiacos son incapaces de responder a todos los impulsos eléctricos, lo que da como resultado la actividad compleja, aparentemente aleatoria y desorganizada, denominada “fibrilación” [21]. Esto provoca que en la señal ECG aparezcan elementos característicos como la ausencia de onda P, una línea base ruidosa o intervalos R-R irregulares (ver Figura 5 y Figura 7).

Actualmente existen dos grandes teorías que intentan explicar el fenómeno de la FA. La primera afirma que es un proceso caótico y autosostenido compuesto por multitud de frentes de onda [22]. La segunda teoría, conocida como la de los rotores madre, afirma que existen fuentes de activación regulares y periódicas durante la fibrilación debido a las reentradas de ondas espirales o rotores en un proceso con cierto grado de organización [23] [24] [25]. En los últimos años esta última teoría es la que está siendo más ampliamente aceptada.

Entre las terapias para la terminación de la FA se encuentra la aplicación de un choque eléctrico (cardioversión eléctrica), o la administración de fármacos antiarrítmicos (cardioversión farmacológica), para así modificar o reiniciar las corrientes eléctricas en el tejido miocárdico con el fin de finalizar el proceso fibrilatorio. Otras terapias más invasivas contemplan la ablación en un procedimiento quirúrgico del tejido miocárdico, normalmente auricular, que provoca la activación caótica del corazón [26] [27].

En cuanto al inicio y progreso de esta arritmia en pacientes, la FA comienza manifestándose en episodios cortos y esporádicos, y con el paso del tiempo aumenta a episodios más largos y frecuentes. Es por ello que se distinguen los siguientes subtipos de FA en la práctica clínica, basándose en su aparición y duración [15]:

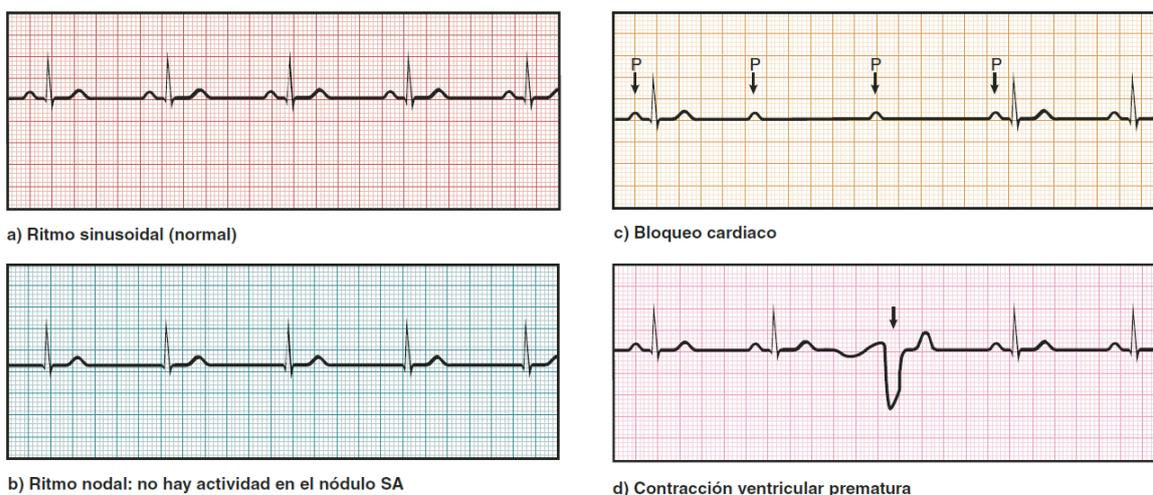
- **FA paroxística.** Episodios que comienzan y finalizan por sí mismos, en periodos no superiores a 48 horas.
- **FA persistente.** Episodios con duraciones superiores a 7 días, incluyendo aquellos que necesiten de cardioversión eléctrica o farmacológica para poder terminar con ellos, aunque no tengan tal duración.
- **FA persistente de larga duración.** Episodios con duraciones superiores a 1 año sin remitir, incluso ante cardioversión eléctrica o farmacológica. Necesitan terapias más agresivas, como la ablación del tejido miocárdico para intentar terminarla.
- **FA permanente o crónica.** Episodios que no remiten ante ninguna terapia.
- **FA valvular.** Producida normalmente por problemas de estenosis (estrechamiento) de la válvula mitral o en pacientes con prótesis valvular mecánica.

## 2.2. Otras arritmias cardiacas

Además de la FA, existen otros tipos de arritmias cardiacas que pueden clasificarse utilizando la siguiente taxonomía:

- Origen anatómico
  - Arritmia Auricular
  - Arritmia Ventricular
  - Arritmia de Unión o Aurículoventricular
- Ritmo Cardíaco
  - Taquicardia (> 100 latidos por minuto en adultos)
  - Bradicardia (< 60 latidos por minuto en adultos).
- Mecanismo arrítmico [28]
  - Automatismo
  - Reentrada
  - Actividad desencadenada
- Conducción Aurículoventricular
  - Normal
  - Retardada
  - Bloqueada
- Duración de la arritmia
  - No sostenida (< 30 s)
  - Sostenida ( $\geq$  30 segundos).

El ECG muestra información de la actividad eléctrica cardíaca, revelando alguna de las arritmias mencionadas. Las diferencias respecto un ritmo normal ayudan a diagnosticar desviaciones en las rutas de conducción, agrandamiento del músculo cardíaco, desequilibrios hormonales o en los canales iónicos celulares, o incluso la aparición de un infarto de miocardio. La Figura 6 muestra algunos ejemplos de estas situaciones.



**Figura 6. Ejemplos de distintos ritmos cardiacos normales y patológicos. a) Ritmo sinusal normal. b) Ritmo generado por el nodo aurículoventricular en ausencia de actividad en el nodo sinusal, provocando ausencia de onda P. c) Bloqueo cardiaco donde algunas ondas P no desencadenan el correspondiente complejo QRS. d) Contracción ventricular prematura o extrasístole [12].**

### 3. Estado del arte

---

Tal y como se ha mencionado anteriormente, la FA se está diagnosticando cada vez más en la práctica clínica. Sin embargo, a pesar del aumento de su prevalencia, esta detección puede ser problemática debido en gran parte a las apariciones episódicas de la arritmia. En relación a esto, el estado del arte de los detectores de FA podría dividirse en las siguientes dos categorías: métodos basados en análisis de la actividad auricular o basados en el análisis de la respuesta ventricular, descritos a continuación.

#### 3.1. Detección de FA basada en la actividad auricular

Los detectores de FA basados en el análisis de actividad auricular se basan en el análisis de la ausencia de ondas P o la presencia de ondas fibrilatorias en el intervalo TQ comúnmente conocidas como ondas *f* (ver Figura 7). Para ello es necesario el estudio y seguimiento de las frecuencias de la FA en el ECG, así como de otros parámetros que intervienen en estas técnicas de procesado de señal.



**Figura 7. ECG de Fibrilación Auricular (arriba) y ritmo sinusal normal (abajo). La flecha roja indica la aparición de ondas *f* en la señal correspondiente a actividad auricular caótica. Además, en el ECG con FA se aprecia un aumento de la variabilidad del intervalo R-R, mientras que en ECG normal de abajo este intervalo es constante. La flecha azulada indica una onda P que se pierde en la fibrilación auricular. En ambos casos, el intervalo QRS indica la activación ventricular. [J. Heuser, Wikimedia Commons, 2005]**

Los métodos publicados para detectar la FA mediante la actividad auricular incluyen: detección de ausencia de la onda P (*P-Wave Amplitude*) [29], análisis del número promedio de ondas *f* [30], aplicación de monitor cardíaco implantable basado en detección de ondas P [31], uso de la entropía *wavelet* [32] [33] y la energía *wavelet* [34], y la utilización de *Echo State Neural Networks* para la cancelación del complejo QRS-T [35].

Para realizar esta tarea, resulta indispensable extraer en primer lugar la actividad auricular de las señales ECG para el análisis de la onda fibrilatoria. Para ello se han de cancelar las ondas prominentemente ventriculares, es decir, cancelar el complejo QRST.

Esta extracción de la actividad auricular durante la FA requiere un procesado de señal no lineal, ya que los espectros de la actividad auricular y ventricular se superponen y, por lo tanto, no pueden separarse por filtrado lineal simple [36]. Existen diferentes estrategias para llevar a cabo esta tarea. Desde los algoritmos más clásicos a los más actuales, nos

encontramos en primer lugar con *Template Matching and Substraction* (TMS), donde se asume que, para cada paciente, las ondas ventriculares muestran un número limitado de morfologías. Un patrón medio de estas morfologías es utilizado para cancelar la actividad ventricular de la señal de ECG [37]. El siguiente método clásico de extracción de señal auricular del ECG es *Blind Source Separation* (BSS), basado en la idea de que la actividad auricular y ventricular se origina a partir de fuentes bioeléctricas diferentes y desacopladas. Mediante el análisis de componentes principales (PCA) [38] [39] [40], o mediante el análisis de componentes independientes (ICA) [36], se buscan las diferentes fuentes no correlacionadas dentro de un ECG, ya sea en una o en múltiples derivaciones. Por otro lado, existen algoritmos basados en redes neuronales y filtrado de Wiener [41], redes neuronales con función de base radial (RBF) [42], o métodos basados en *Least Mean Squares* (LMS) [43] que también han sido utilizados para la extracción de la señal de FA. Finalmente, cabe destacar los métodos basados en *Echo State Neural Networks* (ESN) [35], donde se utilizan paradigmas de entrenamiento de redes neuronales recurrentes basados en *Reservoir Computing*, consiguiendo mejoras en la calidad de la señal de FA extraída respecto de los métodos clásicos. No obstante, todos estos métodos tienen requisitos diversos, tales como la utilización de una o más derivaciones, y presentan diversas variantes que los hacen recomendables o no dependiendo de las distintas situaciones y aplicaciones dentro de la práctica clínica.

Una vez estimada la onda auricular es necesario realizar el análisis espectral de la señal de FA. Para series temporales de electrogramas auriculares es posible obtener la densidad espectral de potencia (PSD), e incluso detectar la frecuencia dominante (FD), que a su vez sirve como marcador de periodos refractarios auriculares y, consecuentemente, de la complejidad de la FA.

Al mismo tiempo, la aplicación de herramientas de análisis tiempo-frecuencia permite revelar la variación temporal de la FD, intentando determinar si esta variación es espontánea o debida a una intervención [44]. En particular, el análisis tiempo-frecuencia se puede utilizar para cuantificar cambios en la frecuencia fibrilatoria y, por consiguiente, su evolución y comportamiento ante medidas terapéuticas externas [26]. Las distribuciones de Choi-Williams o de Wigner-Ville son interesantes y han sido utilizadas en este sentido [45]. Además, cabe destacar que el seguimiento de la evolución de la FD durante la FA puede mejorar su robustez frente al ruido y artefactos utilizando otras técnicas, como es el caso de Modelos Ocultos de Markov (HMM) [46].

Los detectores de FA basados en el análisis de actividad auricular pueden lograr una alta precisión si las señales de ECG registradas tienen poco ruido y alta resolución, pero tienden a sufrir desproporcionadamente en caso de incluir ruido de baja amplitud, el cual se acopla a la señal de FA [47]. Además, cabe destacar que los métodos presentados en este apartado requieren de un alto coste computacional, y normalmente de más de una derivación ECG, con lo que difícilmente pueden ser implementados en dispositivos móviles o *wearables*. Es por ello que las características que serán extraídas de las señales ECG durante este trabajo no contemplarán aquellas relacionadas con la actividad auricular.

### 3.2. Detección de FA basada en la actividad ventricular

Por otro lado, el análisis de la actividad ventricular se basa principalmente en estimar la variabilidad de los intervalos entre latidos cardiacos consecutivos (intervalos RR) utilizando para ello los complejos QRS sucesivos en el ECG. Los intervalos RR se derivan de la característica de mayor amplitud del ECG, cuya detección puede ser mucho más resistente al ruido. En la literatura se considera este enfoque el más adecuado para la detección automática en tiempo real de FA [48].

Los métodos publicados incluyen: análisis del gráfico de dispersión de Poincaré [49] (ver Figura 8), análisis del gráfico de Lorenz [50] [51], análisis de las funciones de distribución acumulativa [52], umbral de la desviación absoluta mediana (MAD) de los intervalos RR [53], primera derivada del histograma de intervalos RR [54], mínimo de la entropía condicional corregida de la secuencia de los intervalos RR [55], detección de irregularidades de los intervalos RR mediante ventana deslizante de 8 latidos consecutivos [56], dinámica simbólica y entropía de Shannon [57], entropía muestral de los intervalos RR [58] [59] [60], y entropía difusa normalizada de los intervalos RR [61].

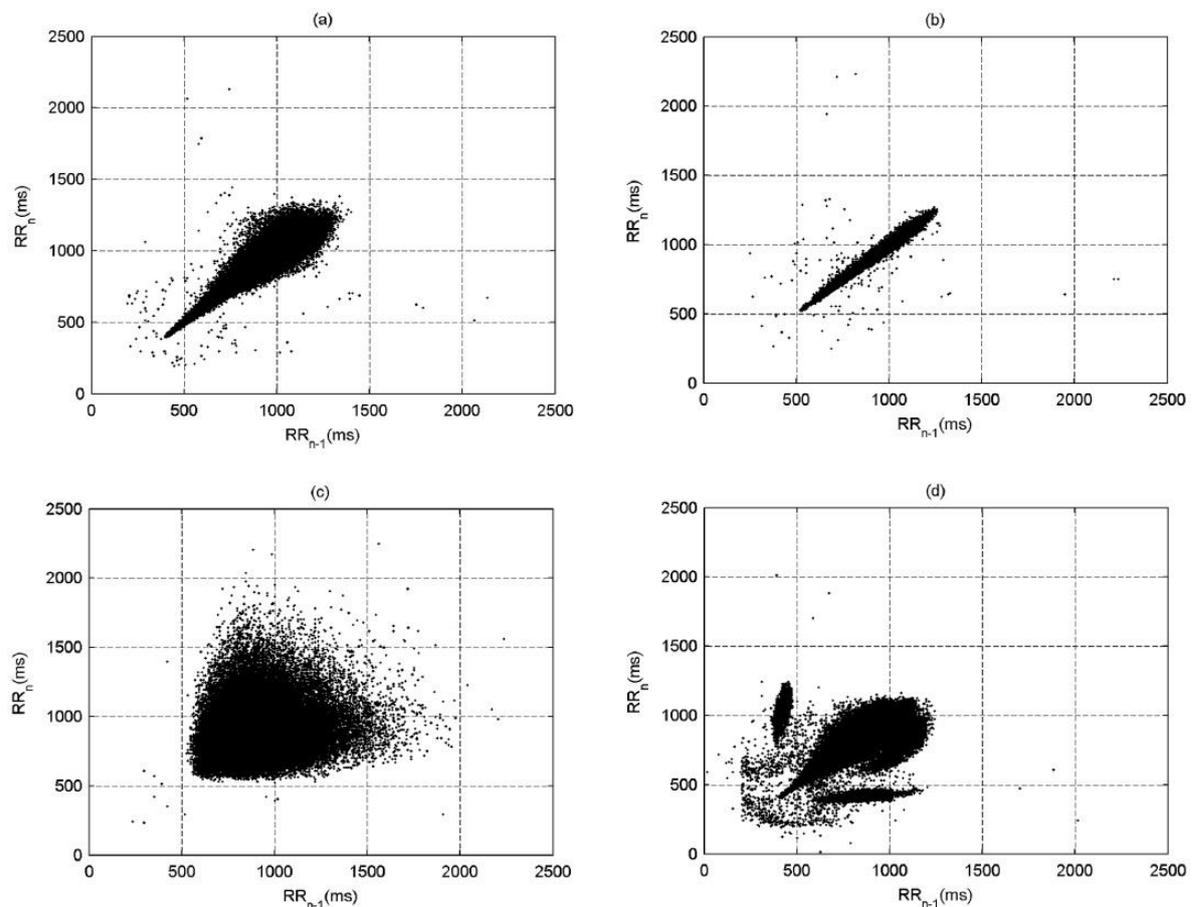


Figura 8. Gráfico de Poincaré, donde cada punto muestra los intervalos de tiempo entre tres latidos consecutivos ( $RR_n$  vs  $RR_{n-1}$ ). El patrón en (b) muestra una variabilidad del ritmo normal, mientras que (a), (c) y (d) muestran distintas arritmias del intervalo RR. Las señales utilizadas contenían 24 horas de monitorización, lo que propicia el elevado número de puntos en las gráficas [51].

Cabe señalar que los detectores de FA 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 cardiaco. Uno de estos enfoques de detección se basó en modelos ocultos de Markov que combinan la variabilidad de los intervalos RR y PR junto con una medida de similitud morfológica de la onda P [62]. Otro de estos estudios aplicó un método de clasificación de lógica difusa que utiliza la combinación de la irregularidad del intervalo RR, la ausencia de onda P, la presencia de ondas *f* y el nivel de ruido [63]. También vale la pena señalar que los enfoques multivariantes basados en el aprendizaje automático que combinan varias de las características individuales anteriores pueden proporcionar una mejor detección de la FA [64] [65] [66].

Finalmente, es necesario indicar que los estudios anteriores sobre la clasificación de FA en general son limitados en cuanto a su aplicabilidad debido a que se realizaron clasificando únicamente ritmos normales y FA. Además, gran parte de estos estudios se realizaron con ECGs de más de una derivación (12 en la mayoría de los casos) y de larga duración, lo que permite la cancelación de ruido y artefactos de una forma más sencilla y fiable. Asimismo, el buen rendimiento que mostraron se debe en parte a que las señales utilizadas fueron seleccionadas cuidadosamente, conteniendo frecuencias limpias y en estudios clínicos con un número reducido de pacientes.

Sin embargo, todavía es un desafío detectar de forma fiable la FA a partir de una única derivación corta del ECG, donde una variedad amplia de los posibles ritmos a encontrar hace esta tarea especialmente difícil. En particular, muchos ritmos anómalos que no son FA presentan intervalos RR irregulares que pueden ser similares a la FA.

Nuestra metodología de clasificación de FA en el contexto del *Challenge 2017 PhysioNet/CinC* [10] aborda este problema principalmente con las aproximaciones basadas en el análisis de la actividad ventricular. Esta decisión se fundamenta en que, al disponer de una única derivación, la señal que aporta la componente auricular normalmente está contaminada por ruido o artefactos que no pueden cancelarse o corregirse con otras derivaciones extra, como sí que sucede en un ECG clásico de 12 derivaciones. Además, el coste computacional de detectar las ondas ventriculares (complejo QRS) es mucho menor que el necesario para el análisis de la actividad auricular descrito anteriormente. Esto es especialmente importante si una de las principales aplicaciones futuras de este trabajo es la posible implementación de los modelos resultantes en dispositivos con pocos recursos computacionales, como son los dispositivos móviles o *wearables*.

### 3.3. Body Surface Potential Mapping

Aunque queda fuera del alcance de este trabajo, es necesario presentar otros sistemas de análisis de la actividad eléctrica cardiaca más avanzados que están desarrollándose actualmente. Estas técnicas están lejos de ser aplicadas como *wearables*. Más bien están pensadas para ser utilizadas en procedimientos quirúrgicos, o análisis clínicos donde se necesita mucha más información de la que pueden aportar los ECGs convencionales de 12 derivaciones o menos.

En primer lugar, existen los registros no invasivos de alta densidad, conocidos comúnmente como *Body Surface Potential Mapping* (BSPM). Estos sistemas utilizan un número elevado de canales o derivaciones, permitiendo obtener de forma más precisa la distribución espacial de la activación eléctrica cardiaca. Mediante estos sistemas se pueden generar mapas de activación eléctrica sobre la superficie del torso, proporcionando información adicional en comparación con los sistemas empleados habitualmente en la práctica clínica, como el ECG de superficie [67] [68].

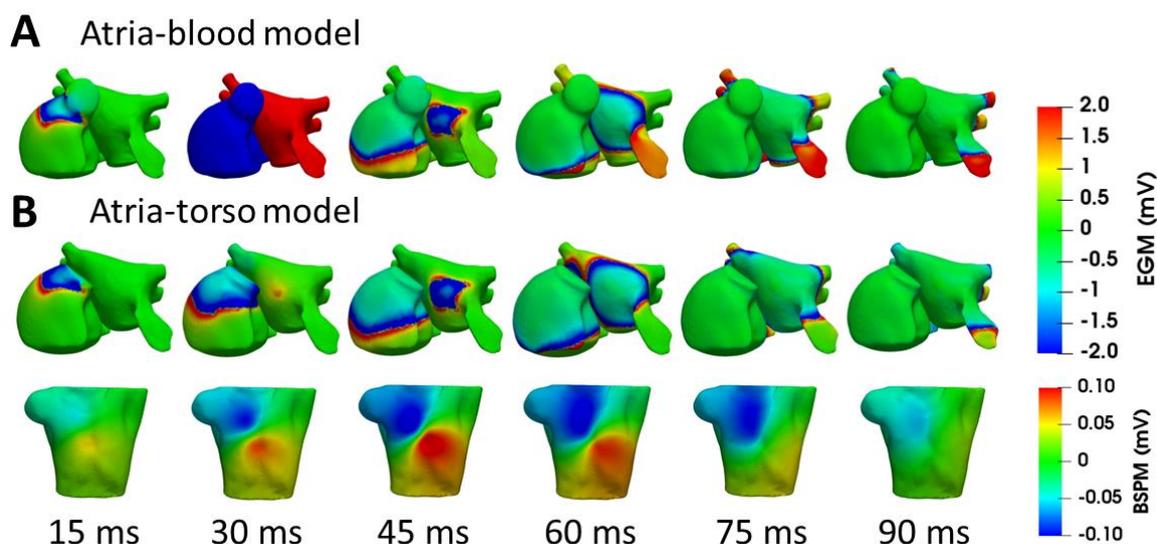
El ejemplo más relevante de BSPM comercial es el sistema *CardioInsight™*, actualmente propiedad de la multinacional *Medtronic*. Este sistema de mapeo 3D no invasivo adquiere 252 señales ECG del torso del paciente (ver Figura 9) y las combina con imágenes de tomografía computarizada (TAC) para producir y visualizar mapas cardiacos tridimensionales [69]. Este mapeo permite al personal médico caracterizar algunos ritmos anómalos del corazón.



Figura 9. Chaleco de adquisición de 252 derivaciones ECG sobre el torso de un paciente del sistema *CardioInsight™*, actualmente propiedad de la multinacional *Medtronic* [69].

### 3.4. Simulaciones computacionales

Finalmente, cabe destacar que mediante modelos computacionales es posible simular una gran cantidad de escenarios relativos a la actividad eléctrica cardiaca. Esto es así gracias al uso de modelos matemáticos que permiten evaluar la actividad eléctrica del corazón a diferentes escalas: canales iónicos, nivel celular y señales registradas en la superficie del cuerpo [70]. Estos modelos típicamente vienen determinados por un gran sistema de ecuaciones diferenciales ordinarias, que son resueltas ayudándose de técnicas de computación de alto rendimiento incluyendo el uso de unidades de procesamiento gráfico (GPU). Gracias al continuo abaratamiento e incremento de la capacidad de cálculo de las GPUs, el número de trabajos de investigación que contemplan simulaciones de la actividad eléctrica cardiaca no ha dejado de aumentar en los últimos años [71]. Estudios recientes han simulado la actividad eléctrica auricular ante diferentes circunstancias, incluyendo distintas fibrilaciones, y posteriormente han comparado y validado los resultados con registros reales intracavitarios obtenidos de sistemas multielectrodo durante procedimientos quirúrgicos [72]. Como resultado, se obtuvieron modelos con los que simular la actividad eléctrica auricular de manera realista (ver Figura 10).



**Figura 10.** Ejemplo de simulación de electrogramas (EGMs) en las aurículas y Body Surface Potential Mapping (BSPM) en un torso. A) Simulación de un modelo de EGMs correspondiente a la interfaz entre aurícula y sangre codificado en escala de colores. B) Simulación de un modelo aurícula-torso que contiene EGMs de la interfaz aurícula-sangre y su BSPM correspondiente sobre el torso. Se muestran instantáneas en distintos instantes de tiempo: 15, 30, 45, 60, 75,-90 ms [72].

Una vez simulada la actividad eléctrica cardiaca y obtenidas las señales correspondientes, se pueden aplicar métodos de análisis de tiempo-frecuencia, entre otros, sobre las señales simuladas. Además, de este modo también es posible obtener patrones correspondientes a diferentes patologías como la FA. Aunque estas simulaciones suponen una simplificación de sistemas biológicos complejos, los modelos matemáticos ya han demostrado su eficacia en la predicción de la respuesta a fármacos [73] [74], por lo que suponen una potente herramienta complementaria en los estudios clínicos y experimentales.

Aunque en este trabajo no se han utilizado estos modelos computacionales, hubiera sido de utilidad la generación sintética de señales ECG de una única derivación con diferentes patologías. De este modo, el número de muestras de nuestra base de datos hubiera aumentado, aunque fuera de forma artificial.



## 4. Marco teórico

---

Bajo la hipótesis comúnmente aceptada de que todos los modelos son erróneos, pero algunos son útiles [75], podemos afirmar que un modelo de clasificación no es más que un intento matemático de representar la realidad. No obstante, estos modelos suelen contener sesgos y errores debido a la complejidad del sistema real a representar. Esto sucede así especialmente en el ámbito clínico, donde los problemas que nos encontramos suelen contener variables desconocidas o difíciles de modelar.

A lo largo de este capítulo se introducirá brevemente los fundamentos teóricos de los distintos tipos de modelos de clasificación que se utilizaron durante la experimentación en este trabajo. Dicha experimentación se compone de un proceso de entrenamiento, validación y evaluación de los clasificadores utilizados, cumpliendo así el ciclo típico dentro del marco del aprendizaje automático y el reconocimiento de patrones.

Así pues, durante la realización de este trabajo se utilizarán tres tipos diferentes de clasificadores: redes neuronales artificiales (RNA), Naïve Bayes y *Support Vector Machines* (SVM), focalizándonos sobre todo en el primero. No obstante, inicialmente se experimentó también con otros clasificadores, como es el caso de los árboles de decisión [76], o *K-nearest neighbours* (K-NN) [77], pero que no han sido validados en este trabajo ya que los resultados preliminares obtenidos no fueron suficientemente buenos.

### 4.1. Aprendizaje Automático y Reconocimiento de Patrones

El aprendizaje automático es un subconjunto de técnicas dentro del ámbito de la inteligencia artificial que, mediante métodos estadísticos, provee a los ordenadores de la capacidad de aprender. Este proceso de aprendizaje consiste en la mejora progresiva del rendimiento en una tarea específica utilizando únicamente datos en forma de observaciones o muestras, suministrando así modelos o estructuras de datos que se adaptan progresivamente a los datos de entrada. Es decir, el resultado del aprendizaje automático no se centra en programar explícitamente los ordenadores para una tarea concreta, sino dotarles de la capacidad de adaptarse y aprender basándose en observaciones o interacciones con el mundo real.

Por otro lado, el campo del reconocimiento de patrones se refiere al hallazgo de características en un conjunto de muestras de datos mediante el uso de algoritmos y procesos informáticos. La finalidad última de la obtención de estas características es la de clasificar las muestras en diferentes categorías [9]. Para ello, se utilizan aquellos modelos matemáticos o estructuras de datos más o menos complejas que consiguen categorizar o etiquetar con mayor precisión una nueva muestra de datos en la clase a la que pertenece. En consecuencia, la generación de modelos para el reconocimiento de patrones requiere de los mecanismos matemáticos y computacionales del aprendizaje automático descritos anteriormente. Gracias a ello, es posible inferir conocimiento de los datos de un dominio concreto.

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 (ver Figura 11). Durante la fase de entrenamiento se utiliza un conjunto de datos para construir el modelo, llamado conjunto de entrenamiento o *training set*. Es en esta fase donde se ajusta 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 incorporarlo en un sistema informático para identificar y clasificar nuevas observaciones. Un ejemplo de estos sistemas serían los SADM que incluyen modelos de clasificación para ayudar a reconocer diferentes patrones clínicos [78].

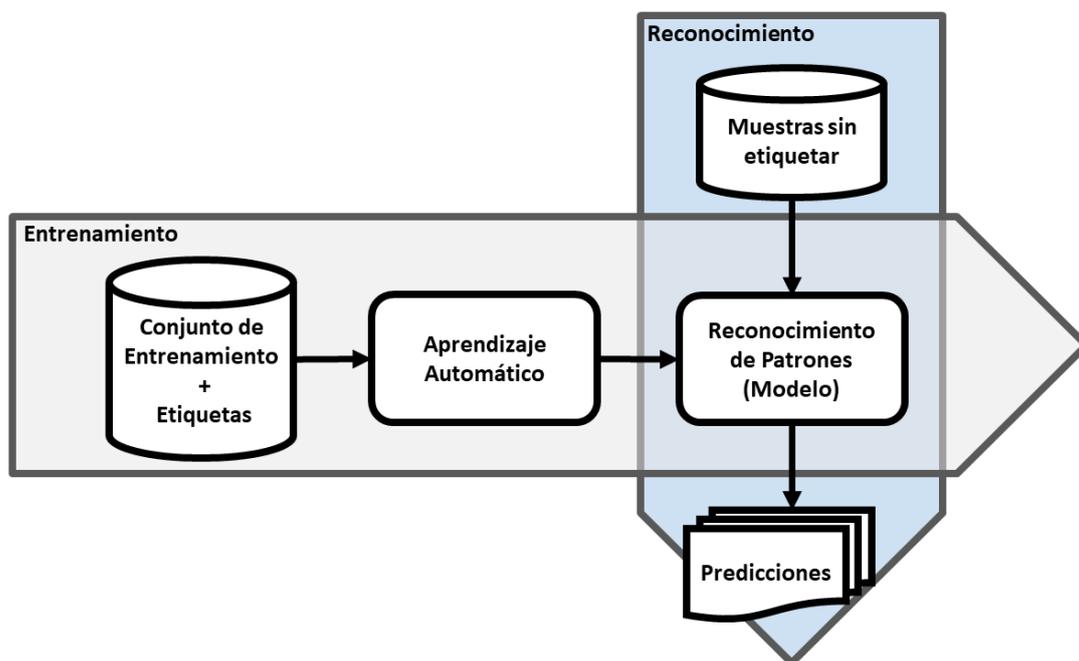


Figura 11. Esquema básico del proceso de generación de un modelo para el reconocimiento de patrones, utilizando un esquema de aprendizaje automático supervisado.

Si se conocen *a priori* las categorías o clases a utilizar se habla de aprendizaje supervisado, ya que el modelo es entrenado para maximizar la precisión de clasificación dentro de ese conjunto de clases. Los modelos se entrenan utilizando muestras ya etiquetadas por personas expertas en la materia con la clase a la que pertenecen. En la última fase, debe ser el modelo generado el que etiquete nuevas muestras sin identificar dentro de las clases conocidas.

Por el contrario, si se desconoce *a priori* la clase o etiqueta de cada una de las muestras se habla de aprendizaje no supervisado, en el que se intenta generar modelos que dividan los datos en agrupaciones ajustadas al conjunto de datos de entrenamiento. El ejemplo más típico de este tipo de aprendizaje automático son los algoritmos de *clustering* o agrupamiento, tales como el k-medias.

Asimismo, el objetivo del aprendizaje automático es desarrollar una regla de decisión o modelo  $M$  que asigne una clase del conjunto  $C$  a un vector de características del conjunto  $X^D$ , donde  $D$  indica la dimensionalidad o número de características que presenta cada una de las muestras [78]. Por lo tanto, un modelo es una asignación con la forma:

$$M: X^D \rightarrow C$$

Un conjunto de entrenamiento viene determinado por los vectores de características utilizados  $x \in X^D$  y sus respectivas clases  $c \in C$ , produciendo muestras  $s_i = \{x_i, c_i\}$  con probabilidad  $p(x_i, c_i)$ , donde  $p(x_i, c_i) = p(x_i)p(c_i|x_i) = p(c_i)p(x_i|c_i)$ . 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.

Generalmente, el modelo se puede definir como una función parametrizada  $\hat{c} = f_M(x, \alpha)$ , que intenta aproximar el valor de  $c$ , y donde  $\alpha$  pertenece al conjunto de parámetros a estimar del modelo. Es posible medir las consecuencias de aproximar  $\hat{c}$  dado  $x$  mediante una función de pérdida (*loss function*)  $L(c, \hat{c})$ , definida normalmente en un problema de clasificación como:

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

Esta función es también conocida como la función de pérdida 0-1. Gracias a ella es posible calcular el riesgo condicional del modelo  $M$ , expresado como:

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

Se define como regla de decisión óptima aquella que consigue una mínima probabilidad de error. Si las probabilidades *a priori* y las probabilidades condicionales de la clase son conocidas, entonces la regla de decisión óptima es la regla de decisión bayesiana:

$$\hat{c} \leftarrow \arg \min_{c \in C} R(c|x)$$

Por otro lado, cuando se asume la función de pérdida 0-1, entonces el riesgo condicional es el error de probabilidad medio, el cual se puede expresar como:

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

Sin embargo, la situación más común es que estas distribuciones de probabilidad sean desconocidas. Aun así, se pueden aproximar mediante el conjunto de observaciones o muestras de entrenamiento  $S = \{(x_1, c_1), (x_2, c_2), \dots, (x_N, c_N)\} \in (X^D \times C)$ , que supuestamente han sido generadas siguiendo  $p(x, c)$ . En un proceso de aprendizaje automático se asume que tanto las muestras conocidas como las no conocidas son generadas por el mismo proceso estocástico. Esto significa que los datos muestreados de manera aleatoria pertenecen a la misma distribución de probabilidad.

El principal objetivo de todo lo descrito anteriormente es encontrar, basándose en el conjunto de muestras conocidas  $S$ , 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:

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

Por otro lado, cuando se asume la función de pérdida 0-1, el riesgo empírico viene determinado por:

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

En la anterior definición del riesgo empírico se hace utiliza la regla delta de Kronecker  $\delta$  definida como:

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

En resumen, durante el proceso de entrenamiento de un modelo para el reconocimiento de patrones, se utilizan algoritmos de aprendizaje automático que ajustan los parámetros del modelo en función de la minimización del riesgo condicional del modelo. Estos algoritmos dependerán del clasificador concreto a entrenar, pero sus fundamentos teóricos son los mismos en todos.

Finalmente, cabe destacar que además de los parámetros del modelo que se ajustan durante el proceso de entrenamiento automáticamente, existen también los hiperparámetros. Podemos definir los hiperparámetros como aquellas variables del modelo que se han de ajustar 'manualmente'. Este ajuste manual o *tuning* se suele hacer midiendo el rendimiento del modelo en un conjunto de datos independiente al de entrenamiento, conocido normalmente por conjunto de validación. Se aplican las mismas reglas de minimización del error del riesgo condicional para entrenar el modelo dentro de un rango permitido de hiperparámetros y se escoge aquel modelo que mejor rendimiento presenta en el conjunto de validación. A este proceso de búsqueda de la mejor combinación de hiperparámetros se le suele llamar red de búsqueda o *grid search*.

## 4.2. Redes Neuronales Artificiales

Las redes neuronales artificiales (RNA) son una familia de modelos inspirados en sistemas biológicos en los que un gran número de unidades simples trabajan en paralelo para realizar una tarea mucho más compleja. Estas redes están compuestas de muchos procesadores simples, también llamados neuronas o unidades, conectados entre sí. Cada una de estas neuronas están basadas en el perceptrón de Rosenblatt [79].

Un perceptrón proporciona una combinación lineal ( $y$ ) de los valores de sus  $n$  entradas ( $x_i$ ), con sus correspondientes pesos ( $w_i$ ), más un valor de *bias* o sesgo ( $b$ ), descrita en la Ecuación 1. La salida del perceptrón se calcula aplicando una función de activación ( $f$ ), a la anterior combinación lineal de pesos y entradas.

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

Ecuación 1. Forma general de la función de activación de un perceptrón con  $n$  entradas.

La Figura 12 muestra un esquema básico del funcionamiento de un perceptrón de estas características. Se ha demostrado que un único perceptrón puede clasificar correctamente muestras que pertenecen a dos clases distintas si son linealmente separables, es decir, es capaz de trazar una frontera mediante un hiperplano para distinguir cada una de las dos clases [80].

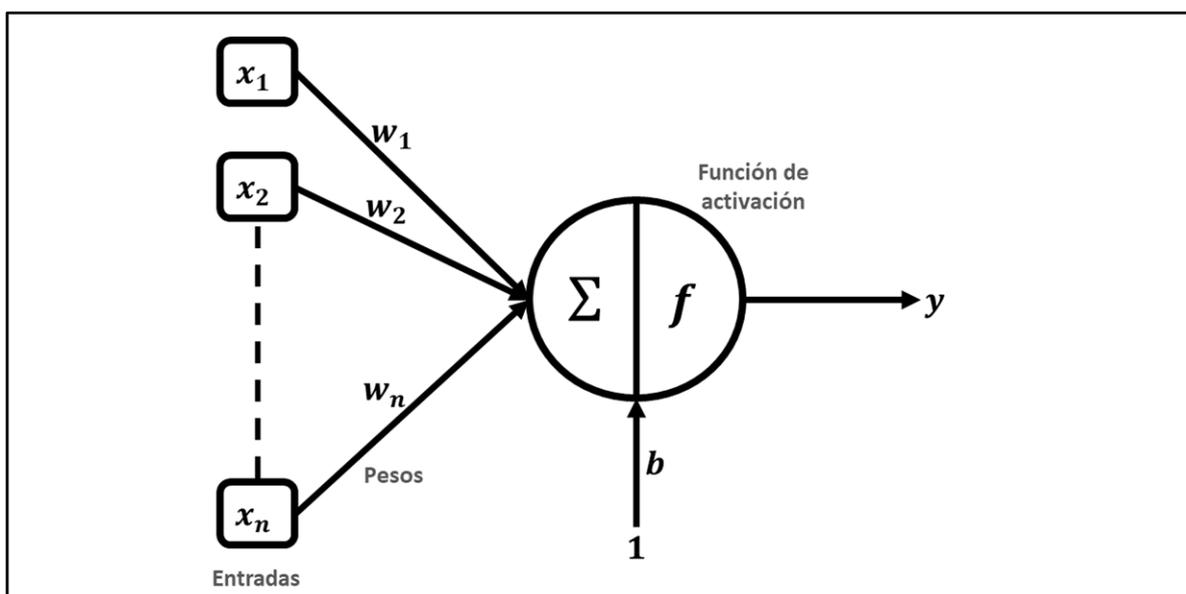


Figura 12. Esquema básico de un perceptrón con  $n$  entradas [81]. A cada una de estas entradas  $x_i$ , se le aplica una combinación lineal de pesos  $w_i$ , y sesgo  $b$ . Finalmente, la función de activación  $f$  determina la salida del perceptrón.

Respecto las funciones de activación utilizadas en un perceptrón, la Tabla 1 muestra una serie de funciones empleadas histórica o actualmente para entrenar redes de perceptrones. Inicialmente se utilizaba la función de paso o *step*. No obstante, puesto que para ajustar los pesos de los perceptrones a partir de muestras etiquetadas se utiliza una optimización de tipo gradiente descendente, la función *step* no resultaba muy adecuada debido a su discontinuidad y derivada nula.

Es por ello que empezaron a utilizarse otro tipo de funciones de activación, como la identidad, tangente hiperbólica o sigmoide. Puesto que estas funciones son monótonicas, la forma de la función de activación de la Ecuación 1 sirve como función discriminante lineal [82].

**Tabla 1.** Lista de algunas funciones de activación utilizadas en perceptrones, junto con sus correspondientes derivadas.

	<b>Función</b>	<b>Derivada</b>
<i>Step</i>	$step(x) = \begin{cases} 1 & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases}$	$step'(x) = 0$
Identidad	$id(x) = x$	$id'(x) = 1$
Tangente Hiperbólica	$tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$	$tanh'(x) = (1 - tanh(x))^2$
Sigmoide	$\sigma(x) = \frac{1}{(1 + e^{-x})}$	$\sigma'(x) = \sigma(x) * (1 - \sigma(x))$
<i>ReLU</i>	$ReLU(x) = max(0, x)$	$ReLU'(x) = \begin{cases} x & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases}$

Asimismo, actualmente también es muy común el uso de otro tipo de funciones de activación, como es el caso de la función de activación ReLU (*Rectified Linear Unit*). Entre las ventajas que se obtienen de utilizar como función de activación la función ReLU, encontramos, en primer lugar, el alivio de carga computacional que supone tanto su cálculo, como el de su derivada, ya que se evita el cómputo de exponenciales o divisiones. Además, se ha demostrado que la convergencia durante el entrenamiento es más rápida gracias a la forma simple de la derivada de la función ante valores positivos, lo que tiene implicaciones positivas a la hora de reestimar los pesos durante el entrenamiento [83]. 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 funciones de activación descritas anteriormente pueden apreciarse en la Figura 13, donde se puede observar las diferencias entre los valores que toman las funciones sigmoide y ReLU, así como su formulación.

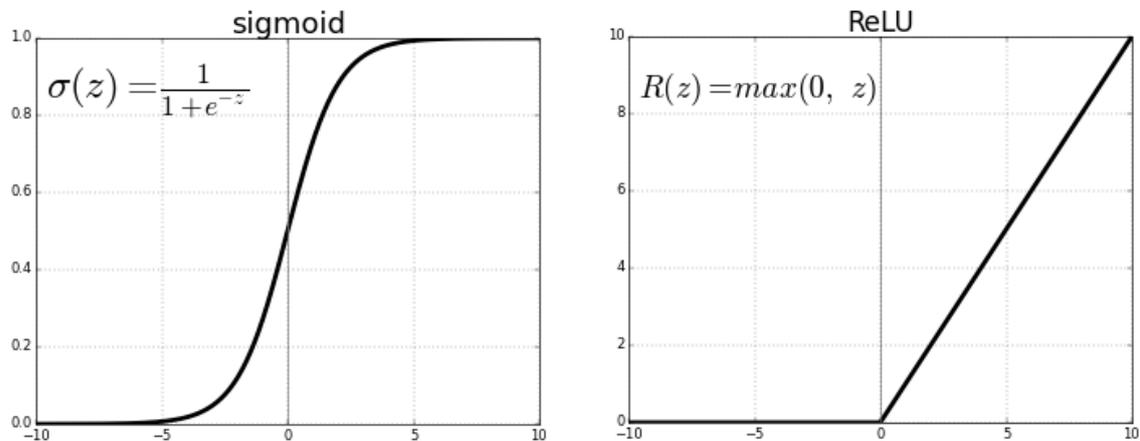


Figura 13. Ejemplos de funciones de activación para un perceptrón: Sigmoide Logística (izquierda) y Rectified Linear Unit, (ReLU), (derecha).

Una vez descritas las características de un único perceptrón, es importante conocer el comportamiento de las redes neuronales compuestas de ellos. Aunque una única unidad tiene una habilidad computacional limitada, un grupo de neuronas interconectadas tienen una capacidad de adaptación muy potente y la habilidad de aprender funciones no lineales. Es así como es posible modelar relaciones complejas entre las entradas y salidas con las que se alimenta a una red de este tipo durante el entrenamiento.

Por lo tanto, se pueden construir funciones más generales al crear redes con sucesivas capas de perceptrones, con conexiones desde cada perceptrón de una capa a todos los perceptrones de la capa siguiente. De ahí la idea del perceptrón multicapa *feed-forward* (FFNN), consistente en una capa de entrada con una unidad para cada variable independiente, seguido de una o dos capas ocultas más de perceptrones. Finalmente, la capa de salida corresponderá a la variable dependiente si estamos modelando una regresión, o, por el contrario, tendremos una unidad por cada una de las clases en la capa de salida si nos encontramos ante un problema de clasificación.

A estas redes, se les llama perceptrón multicapa *feed-forward* totalmente conectado cuando toda unidad de cada capa recibe como entrada la salida de cada unidad de su capa precedente y la salida de cada unidad es enviada a cada unidad de su siguiente capa. La Figura 14 muestra un ejemplo de RNA *feed-forward* con dos capas ocultas y una única unidad de salida.

Se ha demostrado que cualquier zona de decisión arbitraria puede ser aproximada correctamente por una RNA de tipo *feed-forward* con una única capa oculta y un número suficientemente grande de unidades o perceptrones. Para ello es necesario utilizar cualquier función de activación continua como las descritas anteriormente [84].

Estos modelos muestran una gran potencia predictiva comparados con otras aproximaciones tradicionales; sin embargo, se les consideran métodos de 'caja negra'.

Esto es así debido a la falta de capacidad descriptiva sobre la influencia relativa de cada variable independiente en el proceso predictivo. Esta ausencia de poder explicativo es un problema para conseguir una interpretación de la influencia de dichas variables independientes sobre las dependientes.

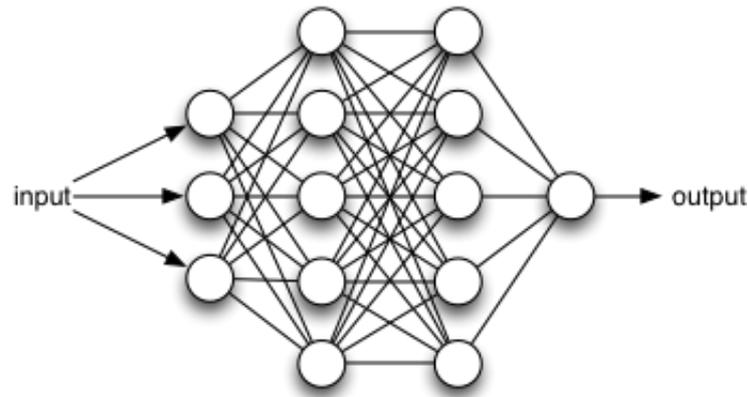


Figura 14: Ejemplo de arquitectura de una red neuronal artificial de tipo *feed-forward*, compuesta de una capa de entrada con tres unidades, dos capas ocultas de cinco unidades cada una, y una capa de salida con una única unidad que representa la función de activación del modelo.

Otra característica que presenta este modelo durante la etapa de entrenamiento es la necesidad de explorar distintas combinaciones respecto al número de capas, número de unidades por capa, e inicialización aleatoria de los pesos para encontrar un modelo que se ajuste bien a los datos de entrada. En nuestro caso, y tal como se explicará posteriormente, se utilizaron perceptrones multicapa *feed-forward* totalmente conectados con una y dos capas ocultas.

### 4.3. Naïve Bayes

El teorema de Bayes intenta estimar la probabilidad de ocurrencia de un evento, basándose en las probabilidades conocidas *a priori* de algunas condiciones asociadas a ese evento. Apoyándose en este teorema y asumiendo independencia entre las variables independientes o predictores surge el clasificador Naïve Bayes [85], el cuál es un modelo fácil de construir y sin ningún hiperparámetro a estimar. A pesar de su simplicidad, en muchas ocasiones muestra un rendimiento sorprendentemente bueno y es ampliamente usado, ya que en algunos problemas mejora los resultados de clasificación obtenidos con métodos mucho más sofisticados.

Dado un conjunto o tupla de predictores  $X$ , donde  $X = \{x_1, x_2, \dots, x_n\}$ , y un conjunto de clases  $C$  tales que  $c \in C$ , el teorema de Bayes proporciona un método para calcular la probabilidad *a posteriori* de la clase  $c$  a la que pertenece un objeto representado por la tupla  $X$  según la Ecuación 2.

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

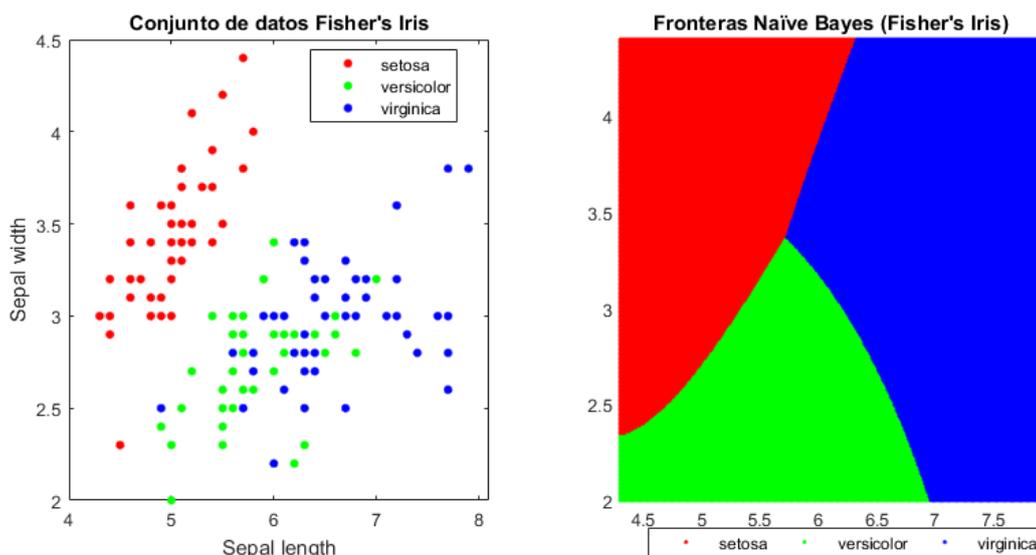
**Ecuación 2:** Descomposición de la probabilidad *a posteriori* de una clase  $c$  con la asunción Naïve Bayes ante un conjunto de predictores  $X$ .

A partir de lo mencionado anteriormente, 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 otro predictor. Esta asunción se llama independencia condicional de la clase y permite que, a la postre, el clasificador se construya basándose únicamente en el numerador del teorema de Bayes. En la Ecuación 3 aparece la regla de decisión básica utilizada para clasificar una nueva muestra como clase  $c$  dados unos datos de entrenamiento (o predictores)  $X$  con Naïve Bayes. Es así como, 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, no existiendo ningún hiperparámetro a estimar.

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

**Ecuación 3:** Regla de decisión en Naïve Bayes para clasificar una muestra como clase  $c$ , dado un conjunto de predictores  $X = \{x_1, x_2, \dots, x_n\}$ .

Las fronteras de decisión que normalmente genera el clasificador Naïve Bayes suelen tener patrones simples y suavizados (ver Figura 15). Aunque estos patrones parecen inicialmente demasiado simples, la generalización que pueden llegar a alcanzar dependiendo del conjunto de entrenamiento es, en muchas ocasiones, bastante buena.

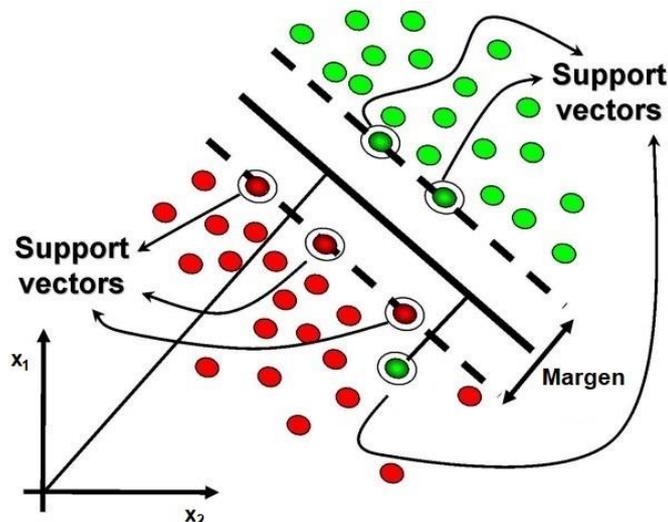


**Figura 15.** Ejemplo de fronteras de decisión generadas mediante Naïve Bayes. Conjunto de datos *Fisher's Iris* ampliamente utilizado en aprendizaje automático (izquierda). Superficies de decisión: setosa, versicolor, virgínica; generadas según el ancho y largo de los sépalos de una flor (derecha).

#### 4.4. Support Vector Machines

En *Support Vector Machines* (SVM) [86], el proceso de aprendizaje automático 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 16 aparece un ejemplo con la idea básica que persigue SVM en dos dimensiones. Contextualizando lo anterior a nuestro caso concreto, y suponiendo que sólo tuviéramos dos variables independientes  $x_1$  y  $x_2$ , los puntos verdes podrían representar personas cuyo ECG muestra un ritmo cardiaco normal, mientras que los puntos rojos aquéllas que padecen fibrilación auricular u otro ritmo cardiaco anormal.



**Figura 16:** Idea básica de hiperplano en SVM en un espacio de dos dimensiones o características  $(x_1, x_2)$ . 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. Si dispusiéramos, por ejemplo, de 70 variables independientes, nuestro espacio estaría compuesto por 70 dimensiones. El hiperplano será de dimensión  $D - 1$ , donde  $D$  indica el número de dimensiones o variables independientes.

Tal y como se observa en la Figura 17, los hiperplanos correspondientes a  $\vec{w} \cdot \vec{x} + b = -1$  y  $\vec{w} \cdot \vec{x} + b = 1$  son los hiperplanos frontera y definen el margen. La distancia entre los dos hiperplanos frontera es el margen, cuyo valor viene determinado por  $2/\|\vec{w}\|$ .

Se puede demostrar que, dados unos datos de entrenamiento, maximizar el margen de separación entre dos clases tiene el efecto de reducir la complejidad del clasificador y, por lo tanto, optimizar la generalización. Es por ello que el hiperplano óptimo corresponde a

aquel que minimiza el error de entrenamiento y, al mismo tiempo, tiene el máximo margen de separación entre las dos clases.

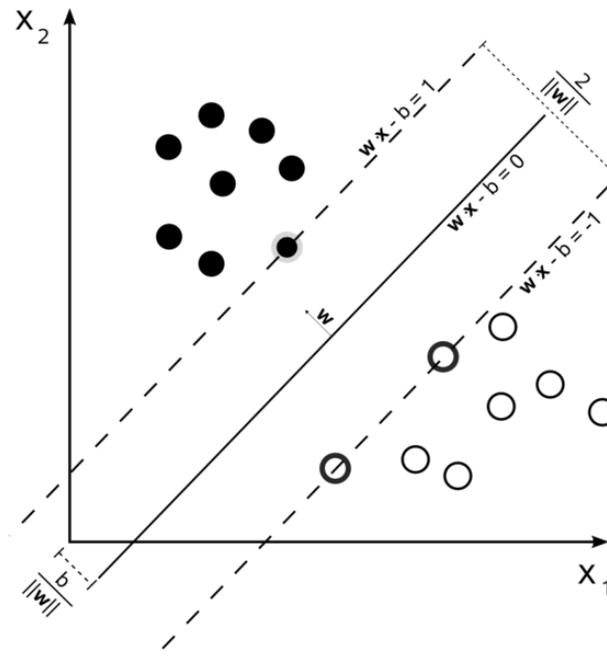


Figura 17: Ejemplo de hiperplano de separación máxima y margen entre dos clases en SVM, en un plano con dos dimensiones  $(x_1, x_2)$ . [Wikimedia Commons, 2018].

Para generalizar los casos donde los límites de decisión no son linealmente separables, SVM proyecta los datos de entrenamiento en otro espacio de dimensionalidad más alta. Si la dimensionalidad del nuevo espacio es suficientemente alta, los datos siempre serán linealmente separables.

Para evitar tener que realizar una proyección explícita en un espacio dimensional mayor se utiliza una función *kernel* ( $K$ ). Esta función  $K$  es la que transforma implícitamente los datos a este espacio dimensional mayor para hacer posible la separación lineal de las clases (ver Figura 18).

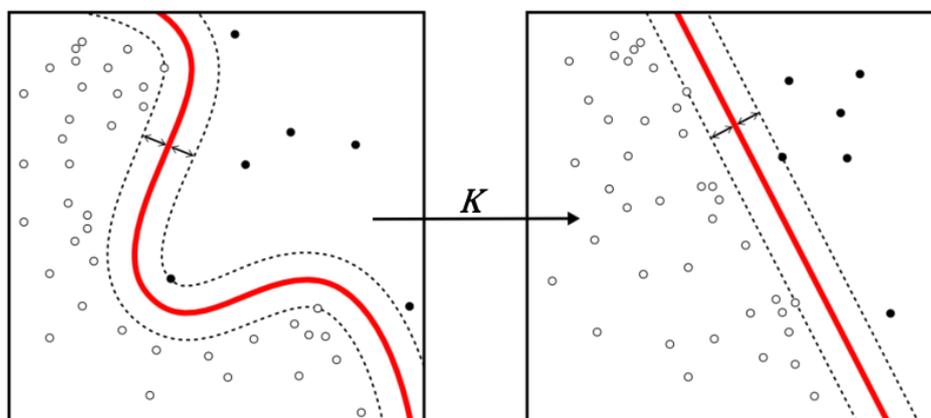


Figura 18. Ilustración de cómo se utilizan las funciones Kernel ( $K$ ) en SVM para transformar un espacio no linealmente separable en otro de mayor dimensionalidad linealmente separable. [Wikimedia Commons, 2018]

La función Kernel  $K$  puede ser de tipo polinomial, de base radial gaussiana, o perceptrón sigmoide, entre otras [81]. La Tabla 2 lista las funciones Kernel más utilizadas en SVM.

**Tabla 2. Listado de funciones Kernel utilizadas en los modelos de clasificación SVM, donde  $s_i$  indica los valores de una tupla de un vector soporte, y  $x_j$  los valores de la tupla del objeto a clasificar.**

<b>Kernel</b>	<b>Función</b>
Lineal	$K_{lin}(s_i, x_j) = s_i \cdot x_j$
Cuadrático	$K_{cua}(s_i, x_j) = (s_i \cdot x_j)^2$
Polinomial	$K_{poly}(s_i, x_j) = (s_i \cdot x_j)^q$
Función de base radial gaussiana (rbf)	$K_{rbf}(s_i, x_j) = \exp(-\gamma \ s_i - x_j\ ^2)$
Perceptrón sigmoide	$K_{mlp}(s_i, x_j) = \tanh(\alpha s_i x_j + \beta)$

Una vez estimados los parámetros del modelo SVM (vectores soporte y parámetros de la función Kernel a utilizar), la regla de clasificación empleada para etiquetar nuevas muestras viene determinada por la Ecuación 4.

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

**Ecuación 4. Regla de clasificación binaria utilizada en SVM para un conjunto de  $n$  vectores soporte, donde  $s_i$  indica cada uno de los vectores soporte usados,  $\alpha_i$  los pesos,  $b$  el *bias* o sesgo,  $K$  la función kernel utilizada y  $\hat{c}$  la clase en que el modelo clasifica la muestra  $x$ . Si el valor del signo de esta función es positivo, la muestra se clasifica como miembro del primer grupo; en caso contrario es clasificada como miembro del segundo grupo.**

A pesar de la potencia de clasificación de SVM, uno de sus desventajas es que sólo puede ser utilizado como clasificador binario. Esto es un inconveniente a la hora de tratar problemas de clasificación con más de dos clases o grupos en los que etiquetar las muestras. Para solucionar este problema se han planteado y evaluado diversas soluciones [87], donde destacan dos aproximaciones: “one-vs-all” y “one-vs-one” [9].

En la primera estrategia, “one-vs-all” o “uno contra todos”, se entrena un único clasificador por clase, con 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” o “uno contra uno”, para un problema con  $C$  clases, se entrena un total de  $C(C - 1)/2$  clasificadores binarios; cada uno recibe las muestras de un par de clases del conjunto de entrenamiento original, y debe aprender a distinguir estas dos clases. A la hora de clasificar nuevas muestras se aplica un esquema de votación donde se asigna la clase que obtuvo más valores positivos (+1) de entre todos los  $C(C - 1)/2$  clasificadores.

## 5. Materiales

---

Toda la parte experimental de este trabajo se llevó a cabo en un PC con procesador Intel® Core™ i7-4790 con una velocidad de reloj de 3.6 GHz, 16 GB de memoria RAM, bajo el sistema operativo Windows 10 Enterprise de 64 bits.

El software utilizado para leer los registros ECG, extraer y seleccionar las características de la señal, y finalmente crear y validar los modelos de clasificación, fue la versión R2017a de Matlab (*Mathworks*).

Además, para el entrenamiento y validación de las distintas redes neuronales artificiales utilizadas y sus hiperparámetros, se utilizó la *toolbox* externa *DeepLearnToolbox* [88] desarrollada también para la plataforma Matlab. No obstante, se implementaron algunas funcionalidades más dentro de esta biblioteca, como la función de activación ReLU, para expandir su funcionalidad.

Respecto al entrenamiento de los demás tipos de modelos (Naïve Bayes y SVM), se utilizaron las funciones estándar de la plataforma Matlab en la versión que utilizamos.

A continuación, se describirá el origen y características de las señales ECG empleadas durante la experimentación en este trabajo, así como el proceso de etiquetado por expertos de esta base de datos para asignar las correspondientes clases a cada una de las muestras.

### 5.1. Registros ECG

En el contexto del “2017 *PhysioNet/Computing in Cardiology Challenge*”, un total de 12.186 registros ECG de una única derivación fueron donados por la empresa AliveCor para esta competición.

Cada señal ECG de un solo canal fue adquirida por una persona que había comprado una de las tres generaciones del dispositivo Kardia Mobile™, de la empresa AliveCor ([www.alivecor.com](http://www.alivecor.com)).

Cada uno de los sujetos sostuvo o presionó, en teoría, un electrodo en cada mano para crear un ECG equivalente a la derivación I (LA-RA, *Left Arm-Right Arm*, por sus siglas en inglés de brazo izquierdo-brazo derecho). Sin embargo, muchos de los ECG aparecen invertidos en esta base de datos (RA-LA) ya que el dispositivo no requiere que el usuario lo gire en ninguna orientación particular.

Para entender mejor lo descrito anteriormente, la Figura 19 muestra el esquema básico de las direcciones que toman las derivaciones más comunes en electrocardiografía (I, II y III) respecto al paciente. Por otro lado, una ilustración del uso del dispositivo de AliveCor aparece en la Figura 20.

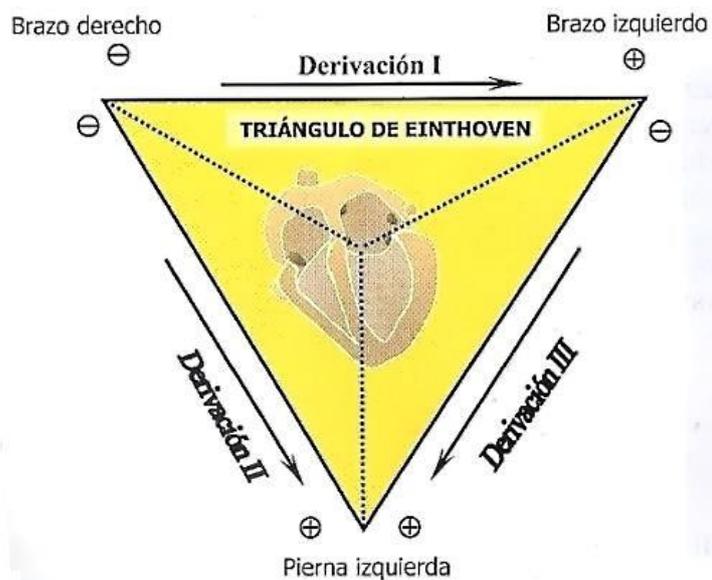


Figura 19. Esquema del Triángulo de Einthoven, en el que se representan las direcciones de cada una de las derivaciones I, II, y III respectivamente respecto al brazo derecho (RA, *right arm*), brazo izquierdo (LA, *left arm*) y la pierna izquierda (LL, *left leg*).

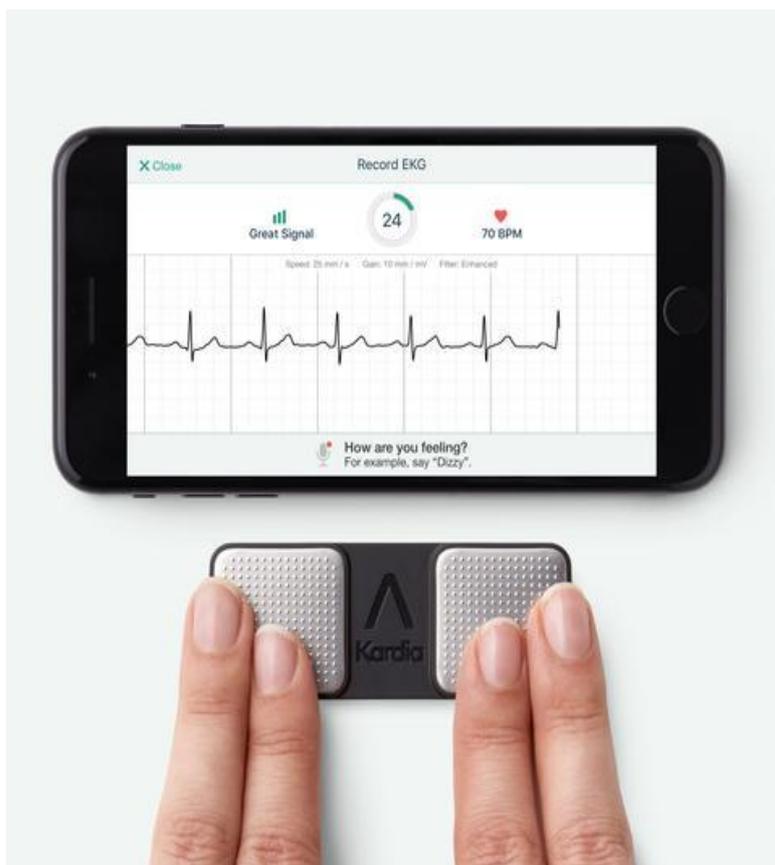
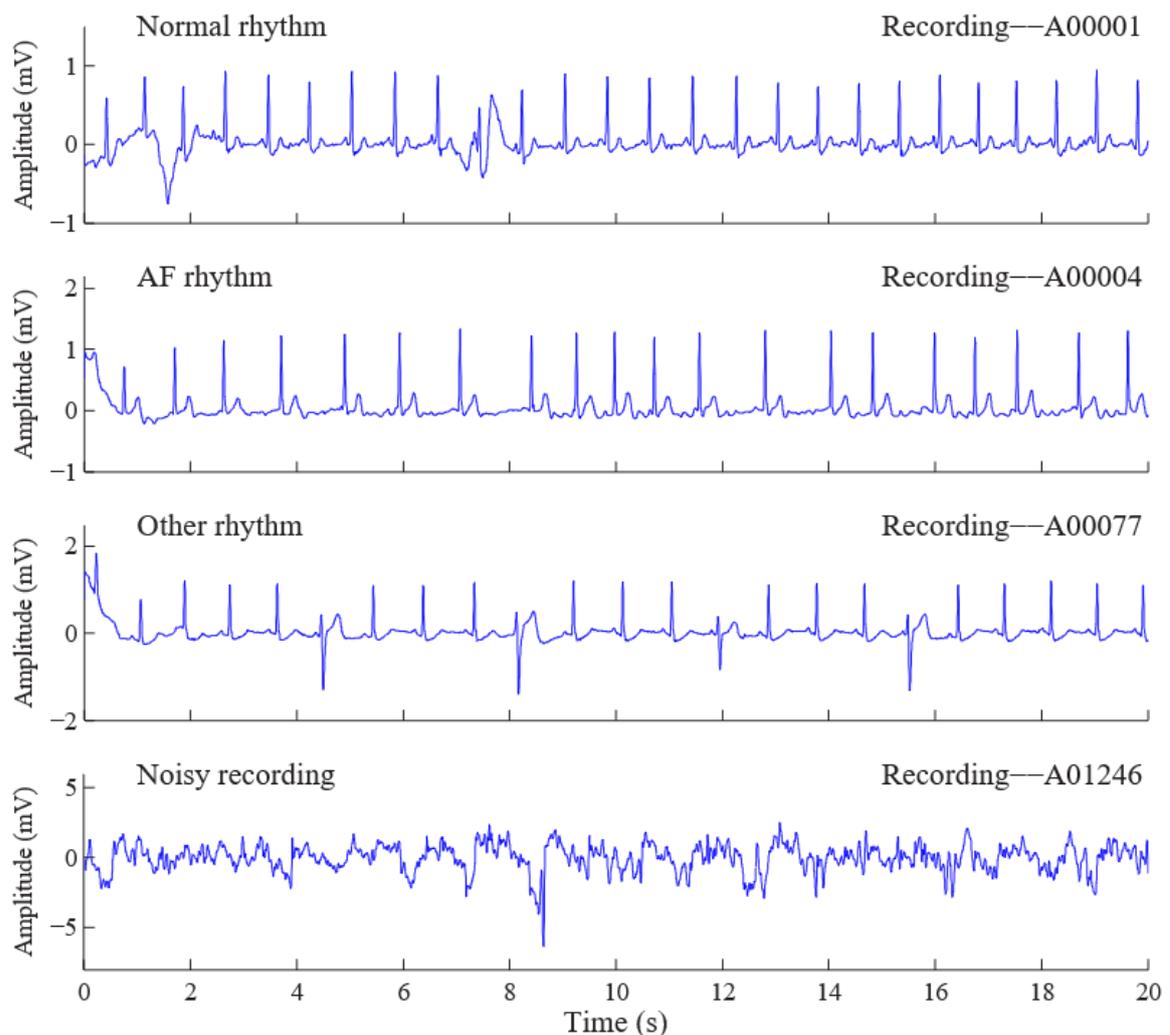


Figura 20. Dispositivo Kardia Mobile™ conectado a la aplicación móvil Kardia App, de la empresa AliveCor, adquiriendo un ECG de una única derivación, equivalente a la derivación I.

Después de algunas comprobaciones básicas de la calidad de la señal, el dispositivo registra el ECG durante un periodo de tiempo promedio de 30 segundos. A continuación, el hardware transmite los datos a un teléfono inteligente o tableta acústicamente, es decir, utilizando el micrófono del dispositivo móvil. Esto permite una comunicación sin cables utilizando una frecuencia portadora de 19 kHz y un índice de modulación de 200 Hz/mV. Seguidamente, los datos son digitalizados en tiempo real a 44.1 kHz y una resolución de 24 bits utilizando demodulación por software. Finalmente, los datos se almacenan en archivos con una frecuencia de muestreo de 300 Hz y una resolución de 16 bits para cada una de las muestras, con un ancho de banda de 0,5-40 Hz y un rango dinámico de  $\pm 5$  mV.



**Figura 21. Ejemplos de registros ECG contenidos en la base de datos utilizada. De arriba a abajo, ECG con ritmo normal, fibrilación auricular, otros ritmos anómalos y registro ruidoso [10].**

En última instancia, los organizadores de la competición convirtieron los datos de los 12.186 registros anteriores en archivos Matlab V4 compatibles con WFDB (*WaveForm DataBase*). Cada muestra constaba de un archivo en formato *mat* que contenía el ECG y otro archivo en formato *hea* que contenía la información de forma de onda. La Figura 21 muestra ejemplos de los ECG descritos anteriormente.

En este trabajo solamente se utilizaron los ficheros en formato *mat*, ya que los *hea* no nos aportaron información útil o fiable.

Para evaluar las distintas metodologías de los participantes en la competición, la base de datos se dividió en dos conjuntos de datos: entrenamiento y test. El conjunto de entrenamiento constaba de 8.528 registros con duraciones desde 9 hasta 61 segundos. Por otro lado, el conjunto de test contenía 3.658 grabaciones de longitudes y distribuciones de clases similares al de entrenamiento. No obstante, el conjunto de test no está a disposición pública y, hasta la fecha del presente trabajo, continúa manteniéndose privado. Esto es así debido a que los organizadores utilizaron el conjunto de test para puntuar a los participantes de la competición y, además, lo reservan durante un período posterior para permitir un seguimiento de las continuaciones de los trabajos de éstos. Estas continuaciones estaban enfocadas sobre todo a la edición del número especial “*Focus on detection of arrhythmia and noise from cardiovascular data*” en la revista científica *Physiological Measurement*.

Aunque este trabajo está basado en la participación del autor en la competición, no todos los clasificadores aquí descritos fueron evaluados con el conjunto de test. Es por ello que se utilizará como base de datos para este estudio únicamente los 8.528 registros de ECG de una única derivación proporcionados públicamente por la competición como conjunto de entrenamiento y validación. No obstante, allá donde sea posible, se indicarán los resultados de clasificación en el conjunto de test.

## 5.2. **Gold Standard** de los registros ECG

Para las agrupaciones de las muestras se consideraron cuatro clases de datos conocidas *a priori*: ritmo normal (N), ritmo de fibrilación auricular (A), otros ritmos (O) y registros ruidosos (-). Durante la competición llegaron a existir tres versiones de las etiquetas de las muestras, aumentando el nivel de precisión en cada una de ellas.

Inicialmente, AliveCor proporcionó las clases de los registros ECG generadas por una tercera empresa. Estas etiquetas se publicaron al comienzo del *challenge* y actuaron como la versión V1 del criterio de referencia o *gold standard*, utilizada en la fase no oficial de la competición.

Sin embargo, algunos registros etiquetados como normales, fibrilación auricular u otros ritmos eran demasiado ruidosos para poder identificar visualmente ningún ritmo. Por lo tanto, los organizadores verificaron todos los ECG visualmente una vez más reetiquetando algunas muestras en la clase ruidosa, dando como resultado la versión V2 del *gold standard*. Este conjunto de etiquetas se utilizó en la fase oficial de la competición.

Finalmente, se creó una tercera y definitiva versión del criterio de referencia para las puntuaciones finales. Para ello, en primer lugar, se seleccionaron las muestras donde los 10 algoritmos con mayor puntuación tuvieron un mayor desacuerdo en su clasificación, obteniendo un total de 1.129 muestras cuya etiqueta no parecía seguir una tendencia clara de clasificación. A continuación, ayudándose de 8 cardiólogos expertos, se revisaron esas 1.129 muestras una vez más para dar lugar a la versión V3 del *gold standard*.

La Tabla 3 muestra la evolución del criterio de referencia durante la competición, siendo la V3 la versión de etiquetado que se puede considerar más precisa. Los resultados de este trabajo corresponden a la versión V3 del etiquetado, a no ser que se indique lo contrario.

**Tabla 3. Evolución del criterio de referencia o *Gold Standard* a lo largo de la competición “2017 PhysioNet/Computing in Cardiology Challenge”, siendo V1, V2 y V3 las tres versiones mejoradas sucesivamente de las etiquetas o clases utilizadas.**

	Número de Registros ECG (%)		
	V1	V2	V3
Normal	5154 (60,4)	5050 (59,2)	5076 (59,5)
Fibrilación Auricular	771 (9,0)	738 (8,7)	758 (8,9)
Otros	2557 (30,0)	2456 (28,8)	2415 (28,3)
Ruidoso	46 (0,5)	284 (3,3)	279 (3,3)

Por otro lado, la Tabla 4 muestra el perfil de duración de los registros ECG acorde a la versión V3 del etiquetado. Recordemos que los resultados de este trabajo serán obtenidos entrenando y validando los modelos con los datos que se hicieron públicos durante la competición, es decir, con 8.528 registros ECG.

**Tabla 4. Perfil de duración de los registros ECG utilizados acorde a la versión V3 del criterio de referencia (media, desviación típica, máximo, mediana y mínimo). Se tiene solo en cuenta las muestras que se hicieron públicas durante la competición.**

	Nº Registros ECG	Duración (segundos)				
		Media	Desv.	Máx.	Mediana	Mín.
Normal	5076	32,1	10,0	61,0	30,0	9,0
Fibrilación Auricular	758	32,3	12,3	60,2	30,0	10,0
Otros	2415	34,3	11,8	60,9	30,0	9,1
Ruidoso	279	24,4	10,4	60,0	30,0	9,4
<b>Total</b>	<b>8528</b>	<b>32,5</b>	<b>10,9</b>	<b>61,0</b>	<b>30,0</b>	<b>9,0</b>



## 6. Métodos

---

Este capítulo describe en primer lugar los criterios de evaluación para los modelos de clasificación utilizados en el presente trabajo. Seguidamente se hablará del preprocesado de señal aplicado a los registros ECG para dejarlos preparados para la extracción de características. A continuación, se listará y explicará cada una de las características extraídas a las señales, algunas diseñadas y desarrolladas durante esta experimentación. A partir del conjunto de datos generado en las etapas anteriores se describirá el proceso de preprocesado de datos y selección de características utilizado. Finalmente se indicará la metodología de validación de los modelos de clasificación y sus hiperparámetros.

Como se ha mencionado anteriormente, todas estas etapas (preprocesado, extracción y selección de características, entrenamiento y validación de modelos de clasificación) se llevaron a cabo utilizando la plataforma Matlab R2017a (*MathWorks*), junto con la *toolbox DeepLearnToolbox* [88].

### 6.1. Criterios de evaluación

A continuación, se explicarán los criterios de evaluación de los modelos de clasificación con los que se experimentó.

Existe una dificultad intrínseca en la naturaleza de nuestro problema puesto que la base de datos utilizada está desequilibrada [89]. Esto significa que los porcentajes de aparición de cada una de las clases varían significativamente: ritmo normal 59,5%, fibrilación auricular 8,9%, otros ritmos 28,3% y registros ruidosos 3,3%. Estas proporciones permitirían que un simple clasificador que asigne la clase más probable *a priori* a una nueva muestra alcanzara una precisión de alrededor del 60%. Sin embargo, su sensibilidad ante muestras con ritmos patológicos sería nula. No obstante, el principal objetivo de este trabajo es obtener modelos predictivos con la mejor sensibilidad posible para cada una de las clases. Es por ello que los criterios de evaluación basados solamente en la tasa de aciertos en una base de datos donde las clases están descompensadas no son un buen criterio para la evaluación de modelos de clasificación.

Para evitar el problema de evaluación descrito anteriormente, los organizadores de la competición escogieron un criterio *F-Score* ( $F_1$ ) para ponderar la precisión y la exhaustividad, más conocidas por sus términos en inglés como *precision and recall* [90].  $F_1$  es una función comprendida entre 0 y 1, alcanzando valores altos sólo si sus dos operandos, precisión y exhaustividad, son elevados y están en equilibrio.

Por lo tanto, la puntuación o *scoring* en este trabajo será la misma que durante las distintas fases de la competición: un valor  $F_1$  promedio de entre las distintas clases. Las reglas de conteo para cada uno de los aciertos y errores de clasificación entre clases se definen en la Tabla 5.

Tabla 5. Reglas de conteo de aciertos y errores entre las distintas clases a etiquetar. La diagonal de la matriz indica los aciertos, donde la clase predicha por el modelo coincide con el *Gold Standard*. El resto de elementos de la matriz representan los fallos, indicando la clase predicha (minúscula) frente a la clase real a la que pertenece la muestra (mayúscula). El número total de muestras por clase y el número total de muestras predichas por clase aparecen en la última columna y fila respectivamente.

		Clase Predicha				
		Normal	FA	Otros	Ruidoso	Total
Gold Standard	Normal	<b><math>Nn</math></b>	$Na$	$No$	$Np$	$\sum N$
	FA	$An$	<b><math>Aa</math></b>	$Ao$	$Ap$	$\sum A$
	Otros	$On$	$Oa$	<b><math>Oo</math></b>	$Op$	$\sum O$
	Ruidoso	$Pn$	$Pa$	$Po$	<b><math>Pp</math></b>	$\sum P$
	Total	$\sum n$	$\sum a$	$\sum o$	$\sum p$	

A partir de las reglas de conteo anteriores,  $F_{1n}$ ,  $F_{1a}$ ,  $F_{1o}$  y  $F_{1p}$  representan la puntuación individual para las clases A, N, O y ~, respectivamente. De este modo, para cada una de las cuatro clases, su valor de *F-Score* se define a continuación.

Ritmo normal:

$$F_{1n} = \frac{2 \times Nn}{\sum N + \sum n}$$

Fibrilación auricular:

$$F_{1a} = \frac{2 \times Aa}{\sum A + \sum a}$$

Otros ritmos:

$$F_{1o} = \frac{2 \times Oo}{\sum O + \sum o}$$

Registro ruidoso:

$$F_{1p} = \frac{2 \times Pp}{\sum P + \sum p}$$

Finalmente,  $F_1$  representa el promedio entre los valores  $F_{1n}$ ,  $F_{1a}$  y  $F_{1o}$ , y en consecuencia el rendimiento de clasificación para un modelo dado en este trabajo.

Nótese que aunque  $F_{1p}$  no se incluye en este promedio, sí que afecta en los valores individuales de  $F$ -Score para cada clase. Por lo tanto, el criterio de evaluación  $F_1$  se define como:

$$F_1 = \frac{F_{1n} + F_{1a} + F_{1o}}{3}$$

Será esta última medida  $F_1$  la que se utilizará para evaluar los clasificadores, ya que relaciona de manera eficaz y comprensible la sensibilidad de detección de cada uno de los ritmos cardiacos. Además, cada uno de los valores individuales  $F$ -Score de cada una de las clases será también presentado en los resultados de este trabajo para intentar comprender así mejor el comportamiento de los modelos obtenidos.

La evaluación final de los modelos se realizó utilizando una validación cruzada de diez particiones con las 8.528 muestras disponibles. Esto es, para un número de particiones  $K$  (diez en nuestro caso), se entrenaron los modelos con todas las combinaciones posibles de  $K - 1$  particiones de muestras, validando el rendimiento en cada ronda con la partición restante ( $F_{1k}$ ). Finalmente, el valor de  $F_1$  no es más que el promedio de estos valores  $F_{1k}$ , definido como:

$$F_1 = \frac{1}{K} \times \sum_{k=1}^K F_{1k}$$

## 6.2. Preprocesado de señal

Las señales proporcionadas por la competición estaban filtradas en la banda entre 0,5 Hz y 40 Hz, por lo que realizamos un filtrado de artefactos como única etapa de preprocesado de señal. Para hacerlo, utilizamos una ventana deslizante de 0,5 segundos para calcular los valores aberrantes máximos y mínimos. Aquellos tramos cuya ventana contenía valores atípicos fueron anulados, tal como aparece en la Figura 22.

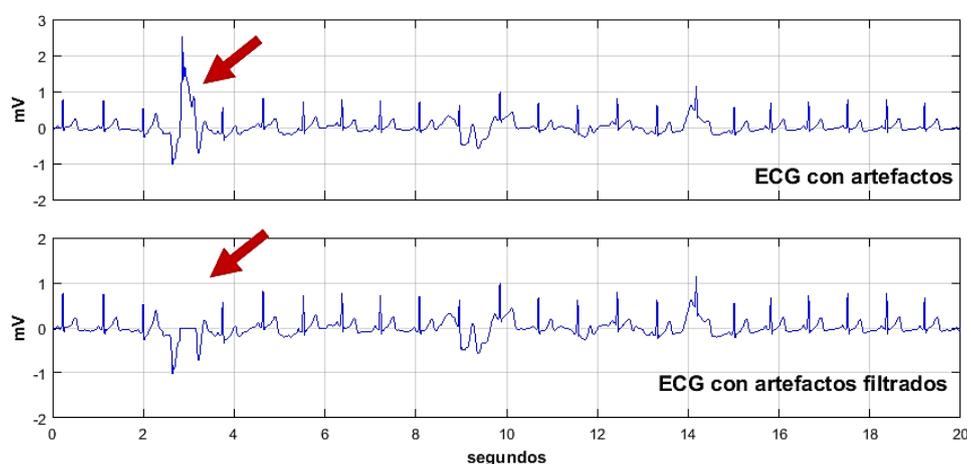


Figura 22. Proceso automático de filtrado de valores aberrantes realizado en las señales ECG.

### 6.3. Extracción de características

A continuación, se presentará una breve descripción de cómo se obtuvieron las variables independientes basadas principalmente en la variabilidad del intervalo RR, utilizadas para desarrollar los modelos de clasificación. Se evitaron las características relacionadas con la actividad auricular, ya que la señal auricular puede verse ocultada fácilmente por artefactos y ruido en señales de una única derivación, o incluso no estar presente por otros motivos.

Se extrajeron automáticamente 81 características de señales derivadas de la actividad ventricular de cada registro de ECG.

Para llevar a cabo esta tarea, inicialmente extrajimos la secuencia RR diseñando y utilizando un detector de ondas R basado en la primera derivada del ECG. 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. De este modo se eliminaron las detecciones en latidos ectópicos y artefactos. La precisión de todas las características extraídas, y por lo tanto del rendimiento de los clasificadores basados en ellas, recae principalmente en la precisión de este detector de ondas R desarrollado para el presente trabajo. La Figura 23 muestra el rendimiento de localización de ondas R de nuestro detector en registros ECG con distintos ritmos cardiacos de la base de datos utilizada.

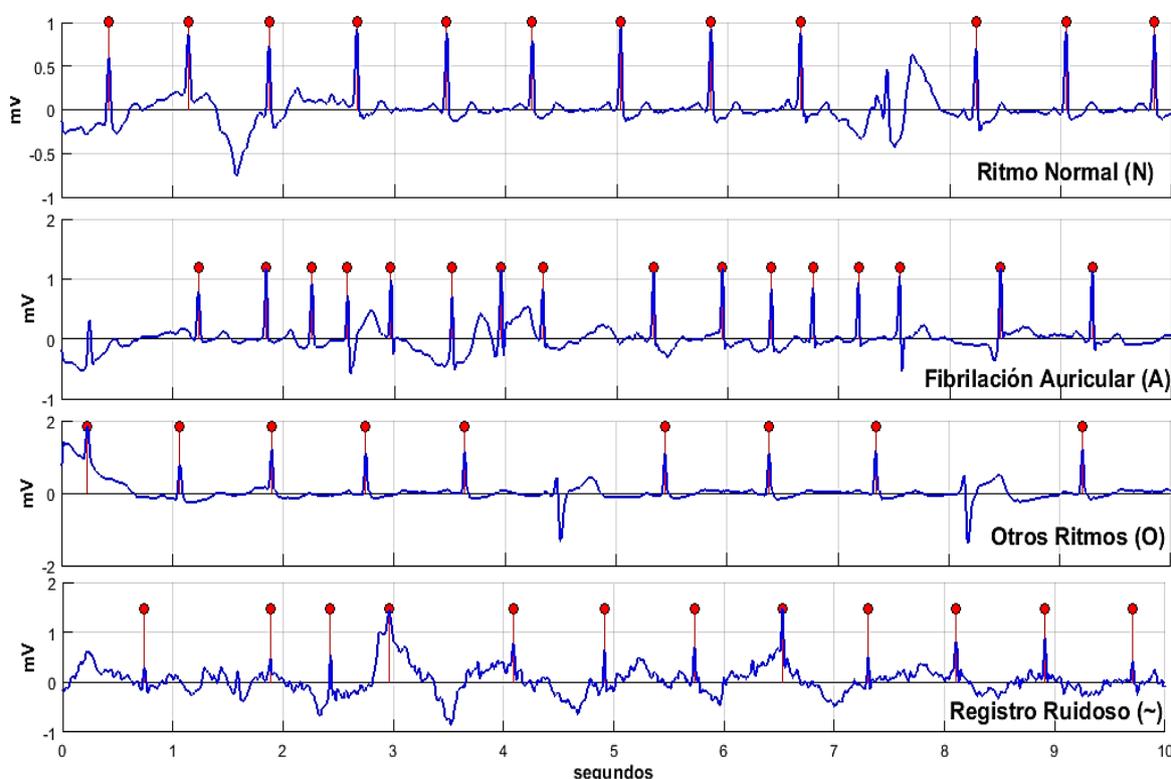


Figura 23. Ilustración del rendimiento del detector de complejos QRS implementado en este trabajo, con registros ECG de la base de datos utilizada. Los puntos rojos indican detecciones de la onda R.

Nótese que la secuencia RR no es más que la derivada de la secuencia temporal de las ondas R detectadas, representando cada uno de sus elementos la diferencia temporal entre dos latidos consecutivos. Cada uno de los  $i$ -ésimos elementos de la secuencia RR vienen definidos por la siguiente expresión:

$$RR_i = R_{i+1} - R_i$$

Una vez obtenida la secuencia RR de cada registro ECG, utilizando como medida temporal los milisegundos, filtramos sus valores atípicos. A continuación, también obtuvimos la primera y segunda derivada de la secuencia RR, a las que llamaremos RRd1 y RRd2 respectivamente, definidas por las siguientes expresiones:

$$RRd1_i = RR_{i+1} - RR_i$$

$$RRd2_i = RRd1_{i+1} - RRd1_i$$

Estas secuencias (RR, RRd1, RRd2) son ampliamente utilizadas para el análisis de la variabilidad del ritmo cardiaco en registros de larga duración (horas o días de señal) [91]. No obstante, nuestros registros contienen duraciones cortas de entre 10 y 60 segundos.

Usando la información anterior, las 81 características que extrajimos de las señales ECG pueden agruparse en los 7 grupos principales que se describen seguidamente. Más información sobre su relevancia clínica puede encontrarse en [92].

Grupo 1. Estadísticas básicas sobre las secuencias RR, RRd1 y RRd2: 9 características.

- Media, desviación estándar, curtosis y asimetría de la secuencia RR.
- Media, desviación estándar y curtosis de la secuencia RRd1.
- Desviación estándar y curtosis de la secuencia RRd2.

Grupo 2. Características basadas en RRd1: 4 características.

- Media cuadrática de diferencias sucesivas (RMSSD).
- Porcentaje de intervalos entre latidos que exceden 25, 50 y 75 milisegundos respectivamente ( $pNN25$ ,  $pNN50$ ,  $pNN75$ ) [93].

Grupo 3. Estadísticas en ventana deslizante. Las mismas estadísticas que en los grupos 1 y 2 (excepto la asimetría del RR, y la media de RRd1) pero usando una ventana deslizante en la señal de 8 segundos en intervalos consecutivos de 2 segundos (11 elementos). Una vez se obtiene la matriz de valores utilizando cada intervalo de señal, extrajimos el mínimo, el máximo, la media y la desviación estándar para cada característica. Todo este conjunto de valores dio como resultado un vector de 44 características.

Grupo 4. Características basadas en el gráfico de Poincaré usando RRd1: 5 características

- Máxima distancia entre dos puntos.
- Mínima distancia entre dos puntos.
- Diferencia absoluta entre la distancia máxima y mínima.
- Media y desviación típica entre todas las distancias.

Grupo 5. Características basadas en el gráfico de Lorenz usando RRd2: 8 características.

- Variabilidad angular: Vector Angular Index (VAI) y Vector Length Index (VLI) [94].
- Media y desviación de las distancias de los puntos al origen.
- Media y desviación de las distancias entre dos puntos consecutivos en RRd2.
- Media y desviación de las distancias entre tres puntos consecutivos en RRd2.

Grupo 6. Características complejas: 3 características.

- Entropía de Shannon de la secuencia RR.
- Complejidad de Lempel-Ziv de la serie temporal RR después de la binarización usando la mediana como umbral.
- Ratio entre el número de distintos patrones QRS encontrados y el número total de ondas R detectadas (PPC).

Grupo 7. Medidas de distancia entre el vector  $pNNx$  medio [93] de cada clase ( $pNNx(A)$ ,  $pNNx(N)$ ,  $pNNx(O)$ ,  $pNNx(-)$ ) y el vector  $pNNx(i)$  correspondiente a cada registro ECG: 8 características.

- Dynamic Time Warping a cada clase ( $dtwNNa$ ,  $dtwNNn$ ,  $dtwNNo$ ,  $dtwNNp$ ) [95].
- Error cuadrático a cada clase ( $ecNNa$ ,  $ecNNn$ ,  $ecNNo$ ,  $ecNNp$ ).

El último grupo de características fue el único que necesitaba un procesamiento previo del conjunto de datos para obtener los vectores  $pNNx$  medios para cada clase. La Figura 24 muestra el resultado de variar el umbral en milisegundos ( $x$ ) para obtener el porcentaje de valores RR consecutivos que difieren más de  $x$  milisegundos ( $pNNx$ ).

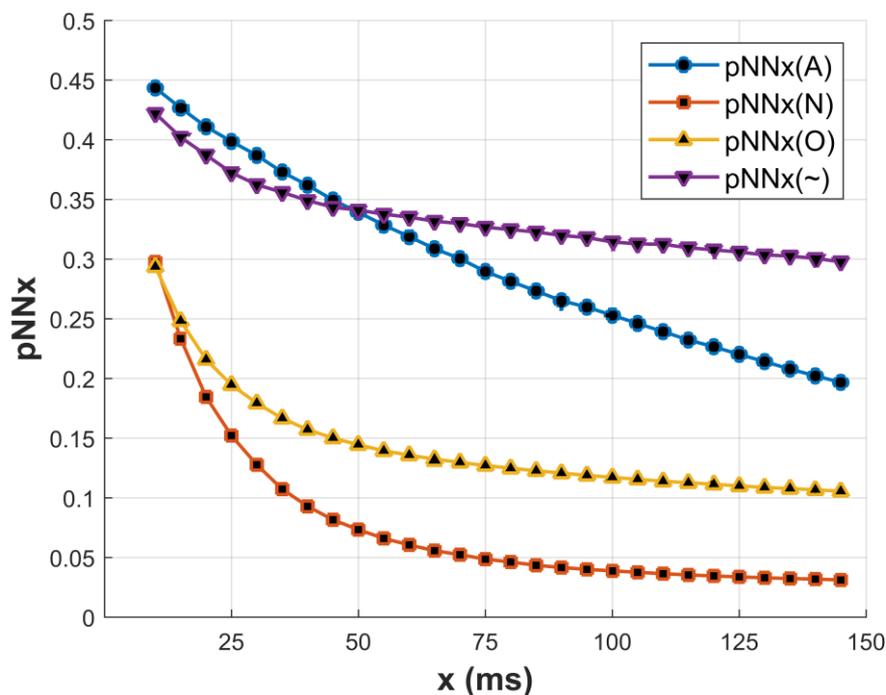


Figura 24. Vectores  $pNNx$  medios obtenidos para cada una de las clases (A, N, O, -) del conjunto de registros ECG: porcentaje de valores RR consecutivos que exceden  $x$  milisegundos.

La Figura 25, además de la media, también muestra la desviación estándar de los vectores  $pNNx$  para las clases N (ritmo normal) y A (fibrilación auricular). Puede apreciarse que la separación de estos dos grupos ayudándose de estos valores puede ser buena. No obstante, el problema principal es el solapamiento que se produce en esta misma representación con la clase O (otros ritmos).

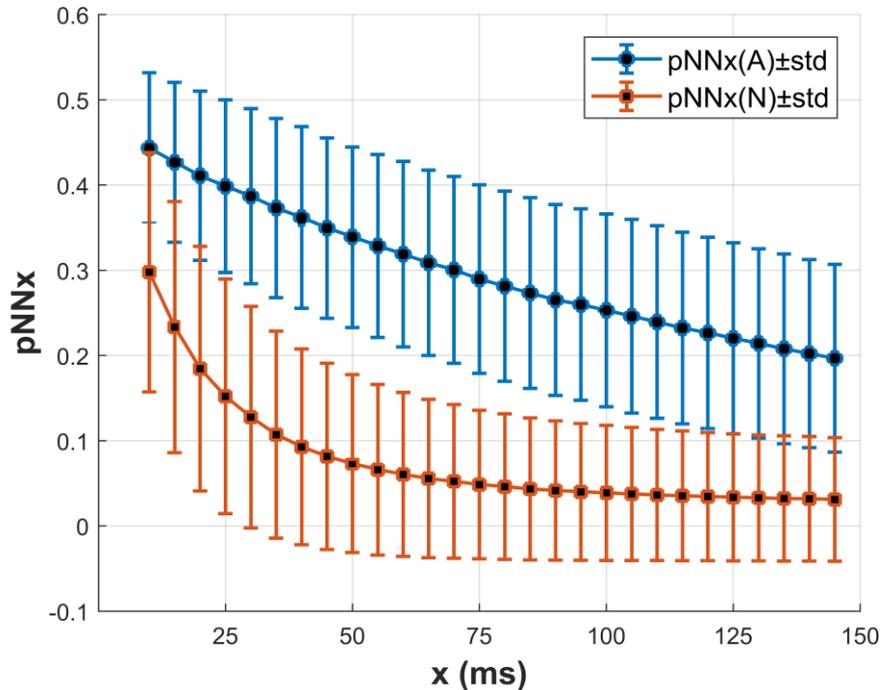


Figura 25. Media y desviación de los vectores  $pNNx$  para las clases A y N.

Usando los vectores presentados arriba, fueron propuestas las ocho nuevas características del grupo 7, basadas en la idea de distancia y similitud entre dos vectores o secuencias. En primer lugar, se utilizó la métrica *Dynamic Time Warping* (DTW) para calcular una distancia entre el vector  $pNNx$  generado para cada muestra  $i$ -ésima ( $pNNx(i)$ ), y cada vector  $pNNx$  medio correspondiente a cada clase:  $pNNx(A)$ ,  $pNNx(N)$ ,  $pNNx(O)$  y  $pNNx(\sim)$ . Estas nuevas características de distancia se definen como:

$$dtwNNA(i) = DTW(pNNx(A), pNNx(i))$$

$$dtwNNn(i) = DTW(pNNx(N), pNNx(i))$$

$$dtwNNO(i) = DTW(pNNx(O), pNNx(i))$$

$$dtwNNp(i) = DTW(pNNx(\sim), pNNx(i))$$

Asimismo, también se utilizó como característica el error cuadrático entre los vectores  $pNNx(i)$  y los cuatro vectores pre calculados  $pNNx(A)$ ,  $pNNx(N)$ ,  $pNNx(O)$  y  $pNNx(\sim)$ , definido como:

$$ecNNA(i) = \sum (pNNx(A) - pNNx(i))^2$$

$$ecNNn(i) = \sum (pNNx(N) - pNNx(i))^2$$

$$ecNNO(i) = \sum (pNNx(O) - pNNx(i))^2$$

$$ecNNp(i) = \sum (pNNx(\sim) - pNNx(i))^2$$

Para ilustrar el comportamiento de las características anteriores, la Figura 26 presenta los diagramas de caja para cada una de las clases utilizando los valores  $dtwNNn$  obtenidos con la base de datos de registros ECG. En ella puede observarse como los valores de distancia  $DTW$  son menores en las muestras que pertenecen a la clase de ritmo normal (N), tal y como cabría esperar.

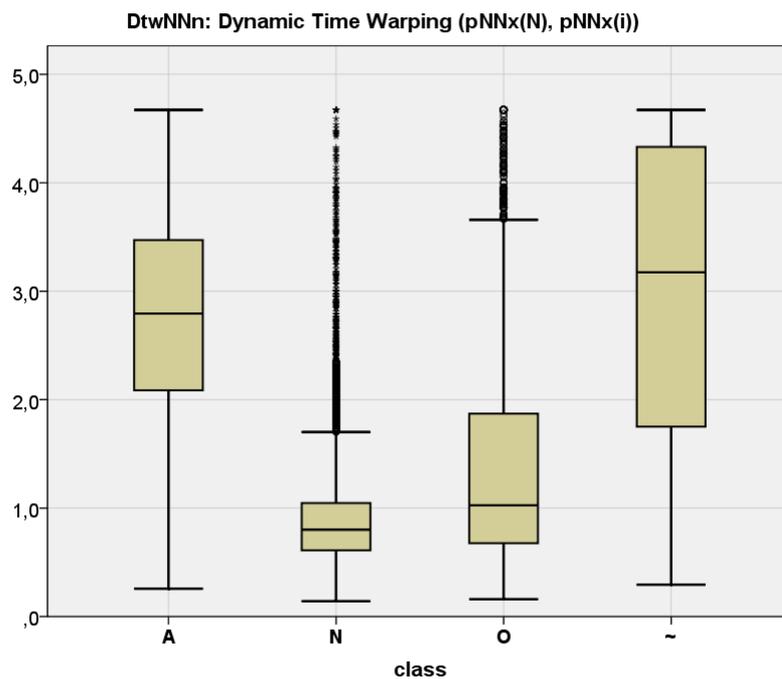


Figura 26. Diagramas de caja de los valores de la característica propuesta  $dtwNNn$ , los cuales representan la distancia  $DTW$  entre cada registro ECG y el vector medio  $pNNx(N)$  correspondiente a la clase de ritmos normales (N).

Utilizando todas las características detalladas anteriormente creamos un conjunto de datos que contenía 8.528 muestras y 81 características para cada muestra.

## 6.4. Preprocesado de las características del conjunto de datos

El objetivo principal del preprocesado de datos es eliminar los posibles errores, así como manipular y transformar los datos en bruto de modo que el contenido de la información pueda ser más fácilmente entendible y manejable [96]. Estos errores pueden ser de muchos tipos: ruido, datos incompletos, duplicados, inconsistencias o valores atípicos. Un valor atípico o *outlier* es aquel que se aleja mucho de la media o la mediana.

El preprocesado del conjunto de datos durante este trabajo constó de tres etapas consecutivas: corrección de valores atípicos, eliminación de muestras con errores y normalización de los datos. A continuación, se detalla cada una de estas fases.

En primer lugar, para cada característica, los valores atípicos que excedían 3 veces la desviación estándar por encima o por debajo de la mediana fueron reemplazados por estos mismos límites.

A continuación, si alguna muestra contenía un valor *NaN* (*Not a Number*) debido a un error de extracción de características, la eliminamos del conjunto de datos. De acuerdo con esta regla, se eliminaron 32 muestras, obteniendo un conjunto de datos final con 8.496 muestras (ver Tabla 6). Puesto que el 56% de las muestras eliminadas correspondían a la clase de registros ruidosos, ante nuevas muestras desconocidas durante la etapa de test en la competición clasificamos como registros ruidosos (~) aquellos que, de entre sus características extraídas, existiera algún valor *NaN*.

Tabla 6. Filtrado de muestras del conjunto de datos por errores en la obtención de características.

	Nº muestras eliminadas	% muestras eliminadas
<b>Total</b>	<b>32</b>	<b>100 %</b>
Ritmo normal (N)	5	15,63 %
Fibrilación auricular (A)	7	21,87%
Otros ritmos (O)	2	6,25 %
Registros ruidosos (~)	18	56,25 %

Nótese que, con 8.496 muestras finales, en la validación cruzada de 10 particiones existe un excedente de seis muestras. Estas seis muestras se incluyeron siempre en el conjunto de validación, siendo el error o acierto que puedan aportar despreciable.

Posteriormente, durante cada ronda de la validación cruzada con  $K$  particiones se obtuvo la media ( $\mu$ ) y la desviación estándar ( $\sigma$ ) de cada característica utilizando únicamente las muestras que pertenecían a las  $K - 1$  particiones utilizadas para realizar el entrenamiento del modelo. Finalmente, aplicamos una normalización *z-score* a las características de todas las muestras, usadas tanto para entrenamiento como para validación. Esta normalización viene determinada por la siguiente expresión, donde  $x$  es el valor de entrada y  $z$  el valor normalizado:

$$z = \frac{x - \mu}{\sigma}$$

## 6.5. Selección de características

Para la selección de características se utilizaron dos aproximaciones distintas: eliminación hacia atrás (*Backward Elimination*) y selección hacia adelante (*Forward Selection*).

En la primera aproximación, *Backward Elimination*, se entrena un modelo con todas las  $N$  características del conjunto de datos y se evalúa su rendimiento inicial. A continuación, se itera entre todas las características eliminándolas del conjunto de datos una a una, entrenando y evaluando el rendimiento del modelo con las  $N - 1$  características restantes. Al final de cada ciclo, se elimina del conjunto de datos definitivamente aquella característica que mayor incremento del rendimiento obtenga, siempre que supere o iguale el rendimiento inicial. Es más, está demostrado que, ante el mismo rendimiento de clasificación, es mejor utilizar aquel clasificador que menos características utiliza, ya que mejora la generalización y reduce la complejidad computacional del clasificador. A continuación se vuelve a repetir el proceso hasta que no se encuentren mejoras en el rendimiento de clasificación.

Respecto a *Forward Selection*, la idea básica es la misma, excepto que el proceso se hace a la inversa: se comienza con un conjunto de características vacío, y se van añadiendo una a una al conjunto de entrenamiento seleccionando en cada iteración aquella que mayor incremento del rendimiento provee. Este proceso se repite hasta que no se observan mejoras en el rendimiento tras añadir alguna característica más. En este caso, igualar el rendimiento de clasificación no sería condición para añadir una característica al conjunto de datos, ya que únicamente se busca que éste incremente con el menor número de características posibles.

Es necesario mencionar que las dos técnicas de selección de características anteriores son de tipo voraz, y por lo tanto subóptimas. Esto quiere decir que el conjunto de características seleccionado finalmente no tiene por qué ser el que proporcione el rendimiento óptimo, pero sí que proporciona un máximo local en el espacio de búsqueda de todas las posibles combinaciones de características.

En este trabajo, el modelo utilizado para realizar los dos procesos de selección de características anteriores fue una red neuronal artificial de tipo *Feed-Forward*. Para evitar el sobreajuste, durante toda la etapa de selección de características utilizamos una arquitectura de red de 1 capa oculta de 128 unidades y solamente 75 *epochs* de entrenamiento, y no las 100 que más adelante se utilizan durante el entrenamiento de los modelos. Además, la partición de los datos utilizada fue una estrategia de retención (*holdout validation*) con el 50% de las muestras destinadas a entrenamiento y el 50% restante a la validación del rendimiento mediante la medida de evaluación  $F_1$ .

Asimismo, se determinó inicialmente de forma empírica los mejores parámetros de esta red neuronal, mediante una *grid-search*, en los valores de velocidad de aprendizaje o *learning rate* ( $\alpha$ ) e impulso o *momentum* ( $\beta$ ), utilizando la estrategia de retención anterior.

Una vez determinados estos parámetros ( $\alpha$  y  $\beta$ ), llevamos a cabo dos procesos distintos de selección de características: el primero utilizando *Backward Elimination* y el segundo mediante *Forward Selection*.

Al proceso de *Backward Elimination* se le añadió una condición de relajación al criterio de parada para evitar una detención anticipada: si durante la iteración actual del algoritmo ( $j$ ) el valor de  $F_1$  no aumenta respecto a la iteración anterior ( $j - 1$ ), seleccionamos como característica a eliminar aquella que maximiza  $F_1$  durante la iteración ( $j$ ) si y solo si  $F_{1j} > (F_{1j-1} - 0.05)$ . En caso contrario, el algoritmo finaliza.

En los resultados de este trabajo se presentarán las características seleccionadas tanto en *Backward Elimination* como en *Forward Selection*, junto con el rendimiento obtenido durante la selección. A continuación, se mostrarán los resultados alcanzados con cada uno de los dos conjuntos de características seleccionadas durante el entrenamiento y validación de los modelos de redes neuronales artificiales, Naïve Bayes y SVM.

## 6.6. Entrenamiento y validación de las Redes Neuronales

Como se ha mencionado más atrás, los hiperparámetros de los distintos modelos de clasificación se seleccionaron empíricamente utilizando una estrategia de validación cruzada con diez particiones.

Utilizando como entradas las características seleccionadas mediante las dos estrategias de selección anteriores (*Backward Elimination* y *Forward Selection*), realizamos una búsqueda de los mejores hiperparámetros de entrenamiento para una red neuronal de tipo *feed-forward*: arquitectura (número de unidades por capa), *learning rate* ( $\alpha$ ) y *momentum* ( $\beta$ ).

Los modelos aquí entrenados pretenden mejorar una experimentación previa realizada para la competición y publicada en [95]. Debido a esto ya eran conocidas algunas condiciones iniciales de los hiperparámetros, como que las redes neuronales con dos capas ocultas no mejoraban el rendimiento de clasificación. Finalmente, los rangos de valores utilizados durante la búsqueda en este trabajo fueron los siguientes:

1. Redes neuronales de una capa oculta con 64, 128, 256, 512 y 1024 unidades respectivamente.
2. Valores de *learning rate* ( $\alpha$ ) de 0.5, 0.6, 0.7, 0.8, 0.9 y 1.0 respectivamente.
3. Valores de *momentum* ( $\beta$ ) de 0.0, 0.1, 0.2, 0.3, 0.4 y 0.5 respectivamente.

Durante el entrenamiento se utilizó optimización *Levenberg-Marquardt backpropagation* por descenso de gradiente para actualizar los pesos de las redes neuronales [97]. Además, como condiciones de parada del aprendizaje se usó un máximo de 100 *epochs* y un incremento mínimo del gradiente de 0.001 entre dos *epochs* consecutivas. Siguiendo esta estrategia, se entrenaron y validaron un total de 180 combinaciones de hiperparámetros en 1800 ciclos de entrenamiento y validación debido a las 10 particiones de datos.

Como función de activación de todas las unidades se utilizó la función ReLU, excepto para las unidades de la última capa de la red en las que se utilizó la función *softmax*, también

conocida como función exponencial normalizada [9]. Además, se utilizó una tasa de *dropout* de 0.5, o lo que es lo mismo, en cada *epoch* se actualizaron solamente la mitad de las unidades de las redes neuronales y sus pesos, ya que se observó empíricamente que se obtenía una mejor generalización con el conjunto de datos utilizado.

## 6.7. Entrenamiento y validación de Naïve Bayes

La experimentación en este trabajo con los modelos Naïve Bayes consistió en el entrenamiento y validación cruzada del rendimiento de clasificación en los dos conjuntos de datos seleccionados anteriormente. En ambos experimentos se utilizó la misma búsqueda en las distintas configuraciones que pueden ser utilizadas con Naïve Bayes.

La primera configuración que puede ser elegida en este tipo de modelos tiene que ver con el tratamiento que se les dé a las probabilidades *a priori* de cada una de las clases. Es posible escoger entre dos posibles valores: (1) empíricas, (2) uniformes. En el primer caso, las probabilidades *a priori* de las clases son determinadas empíricamente del conjunto de datos de entrenamiento; es decir, ante un conjunto de datos de dos clases  $C_1$  y  $C_2$ , en el que las muestras de  $C_1$  representan el 75% de los datos, los valores de probabilidades *a priori* serían de 0.75 y 0.25 respectivamente para cada una de las clases. En el segundo caso, utilizando probabilidades uniformes en un problema con dos clases, los valores de probabilidad para cada una de ellas serían de 0.5 y 0.5 respectivamente. En la práctica, utilizar probabilidades de la clase *a priori* uniformes significa eliminar este término de la ecuación, ya que no afecta en el resultado de la clasificación.

La segunda configuración posible a elegir en un modelo Naïve Bayes tiene que ver con la distribución de los datos utilizada para modelar la probabilidad de cada una de las características del conjunto de datos.

Una de las distribuciones más utilizadas en este sentido es la distribución Gaussiana, definida por una media ( $\mu$ ) y una desviación típica ( $\sigma$ ). Al clasificar una nueva muestra con valor  $x$  en una de sus características, podemos modelar la probabilidad de aparición del valor  $x$  mediante la siguiente función:

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

Otra función utilizada para estimar la distribución de los datos en Naïve Bayes es una función de densidad basada en Kernel (*Kernel Density Function, KDF*) definida como:

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

donde  $K$  es la función Kernel a utilizar,  $n$  el número de muestras del conjunto de entrenamiento y  $W$  un parámetro libre de suavizado llamado típicamente *width*, *bandwidth* o *kernel smoother window width*. En la función anterior también es posible indicar un valor

$W$  para cada una de las clases del conjunto de datos. Además, el valor de  $W$  puede ser tratado como un hiperparámetro a estimar para mejorar el rendimiento del clasificador Naïve Bayes.

Con todo lo descrito anteriormente, en este trabajo se realizó una *grid-search* en los modelos Naïve Bayes en el conjunto de configuraciones siguientes:

1. Probabilidad *a priori* de las clases
  - a. Empírica
  - b. Uniforme
2. Distribución de datos
  - a. Gaussiana
  - b. *Kernel Density Function (KDF)*
3. Para las distribuciones de probabilidad *KDF*, se utilizó todas las posibles combinaciones entre los valores 0.005, 0.020, 0.035 y 0.050 para el *bandwith* ( $W$ ) de cada clase ( $W_n, W_a, W_o, W_p$ ).

El rango de valores  $W$  usado en las funciones *KDF* en este trabajo fue seleccionado empíricamente en experimentaciones previas no descritas aquí por simplicidad, pero que aumentaron el número de pruebas reales ejecutadas.

Por otro lado, la función Kernel utilizada fue una distribución normal con media 0 y desviación estándar 1, definida como:

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

Siguiendo esta estrategia, se entrenaron y validaron un total de 516 combinaciones de hiperparámetros en 5160 ciclos de entrenamiento y validación debido a las 10 particiones de datos, para cada una de los conjuntos de datos seleccionados.

## 6.8. Entrenamiento y validación de SVM

A continuación, se describe la experimentación realizada con los modelos de tipo *Support Vector Machine*, la cual consistió, al igual que en los modelos anteriores, en el entrenamiento y validación cruzada del rendimiento de clasificación en los conjuntos de datos seleccionados. En ambos experimentos se utilizó la misma búsqueda en los distintos hiperparámetros que se detallan seguidamente.

Como se ha descrito anteriormente (ver sección 4.4. *Support Vector Machines*), en SVM se clasifica mediante el hiperplano que maximiza el margen entre dos clases en los datos de entrenamiento. Los vectores que definen este hiperplano, seleccionados de entre los predictores o variables independientes, son los llamados '*support vectors*'.

Los modelos SVM están definidos como clasificadores binarios (dos clases a etiquetar), donde no existe selección de ningún umbral de decisión en su salida, ya que la clasificación se realiza etiquetando cada muestra a un lado u otro del hiperplano creado.

Para convertir SVM en un clasificador de más de dos clases, en este trabajo se siguió la estrategia “one-vs-one” o “uno contra uno”, donde es necesario entrenar un total de  $C(C - 1)/2$  clasificadores binarios en un problema con  $C$  clases. En nuestro caso, con cuatro clases, entrenamos un total de seis clasificadores e implementamos el correspondiente sistema de votación entre las salidas de cada uno de los modelos para clasificar una muestra de nuestra base de datos. En cada punto de la red de búsqueda de hiperparámetros de SVM se utilizaron los mismos valores para entrenar los seis clasificadores binarios, evitando así una explosión combinatoria en la práctica inabordable.

Además, durante la experimentación se utilizaron solamente *kernels* de base radial gaussiana (*rbf*), definidos como

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

donde el parámetro libre gamma ( $\gamma$ ) provee un factor de escala para la función, y las tuplas  $s_i$  y  $x_j$  representan un *vector soporte* y la muestra a clasificar respectivamente.

Por otro lado, con el fin de permitir cierta flexibilidad, los modelos SVM manejan una variable  $C$  que controla la compensación entre errores de entrenamiento y los márgenes rígidos, creando así un margen blando (*soft margin*) que permita algunos errores en la clasificación a la vez que los penaliza [81]. De este modo, el hiperparámetro  $C$  representa el compromiso entre el tamaño del margen y el número de errores de clasificación.

Con todo lo descrito anteriormente, se puede concluir que el rendimiento del SVM depende de la función *kernel* utilizada, sus parámetros y el parámetro de penalización del margen  $C$ .

La búsqueda realizada en este trabajo en los hiperparámetros de los modelos SVM fue la siguiente:

1. Valores de *soft margin* ( $C$ ) en el rango [0.2, 1.0] con incrementos de 0.1 respectivamente (9 valores distintos).
2. Valores del factor de escala en la función *rbf* ( $\gamma$ ) en el rango [1.0, 3.0] con incrementos de 0.1 respectivamente (21 valores distintos).

El rango de valores de los hiperparámetros anteriores fue seleccionado empíricamente en experimentaciones previas no descritas aquí por simplicidad. En total, si contamos todas las combinaciones de hiperparámetros descritas se probaron 189 parametrizaciones con SVM. No obstante, y tal como se describirá en los resultados, se encontró un patrón de rendimiento definido en el *grid* de hiperparámetros, con una zona donde los valores de  $F_1$  se incrementaban. Es por ello que se realizó otra búsqueda más detallada en esa zona, encontrando el modelo que mejor rendimiento de clasificación se consiguió con SVM.

## 7. Resultados

---

En el presente capítulo se presentarán los resultados experimentales obtenidos durante este trabajo. En primer lugar, se mostrarán los resultados de la selección de características. A continuación, se presentará el rendimiento de los modelos de clasificación basados en redes neuronales, Naïve Bayes y SVM. Finalmente se muestra una breve visión global de los resultados anteriores.

### 7.1. Selección de características

En primer lugar, se determinaron empíricamente mediante *holdout validation* los hiperparámetros de la red neuronal de una capa oculta y 128 unidades utilizada para seleccionar las características del conjunto de datos. Este modelo inicial, entrenado utilizando las 81 variables independientes disponibles, mostró un rendimiento en el criterio de evaluación  $F_1 = 0.7357$ , siendo el rendimiento individual en cada una de las clases el siguiente:  $F_{1a} = 0.7641$ ,  $F_{1n} = 0.8328$ ,  $F_{1o} = 0.6096$  y  $F_{1p} = 0.4841$ .

Una vez seleccionado el *learning rate* ( $\alpha$ ) y el *momentum* ( $\beta$ ) a utilizar durante toda la fase de selección de características, se emplearon las estrategias *Backward Elimination* y *Forward Selection* para generar dos conjuntos con menos variables que el inicial. La Tabla 7 muestra los resultados finales obtenidos con cada una de las aproximaciones.

**Tabla 7. Resultados de la selección de características mediante las dos técnicas utilizadas. El número de características indica cuántas sobrevivieron al proceso de selección de entre las 81 que contenía el conjunto de datos original. También aparecen los valores *F-Score* obtenidos en el conjunto de validación (experimentación *holdout validation*).**

	Backward Elimination	Forward Selection
<b>Nº características seleccionadas</b>	<b>18</b>	<b>15</b>
$F_1$	0.735	0.746
$F_{1a}$	0.764	0.787
$F_{1n}$	0.843	0.849
$F_{1o}$	0.599	0.601
$F_{1p}$	0.497	0.503

En *Backward Elimination* se consiguió eliminar 63 características, obteniendo un conjunto final de 18 atributos con el que se igualaba el rendimiento del modelo inicial. Por otro lado, *Forward Selection* eligió un conjunto total de 15 características con el que mejoró ligeramente el rendimiento de los modelos anteriores.

Las dos estrategias anteriores coincidieron en la selección de un total de cinco características: (1) valor máximo del *pNN50* en el conjunto de ventanas deslizantes; (2) valor medio de la desviación de RRd1 en el conjunto de ventanas deslizantes; (3) asimetría de la secuencia RR; (4) proporción entre el número de distintos patrones QRS encontrados

y el número total de ondas R detectadas (PPC); (5) desviación estándar entre todas las distancias de los puntos del gráfico de Poincaré. La Figura 27 muestra el diagrama de cajas de cada una de ellas.

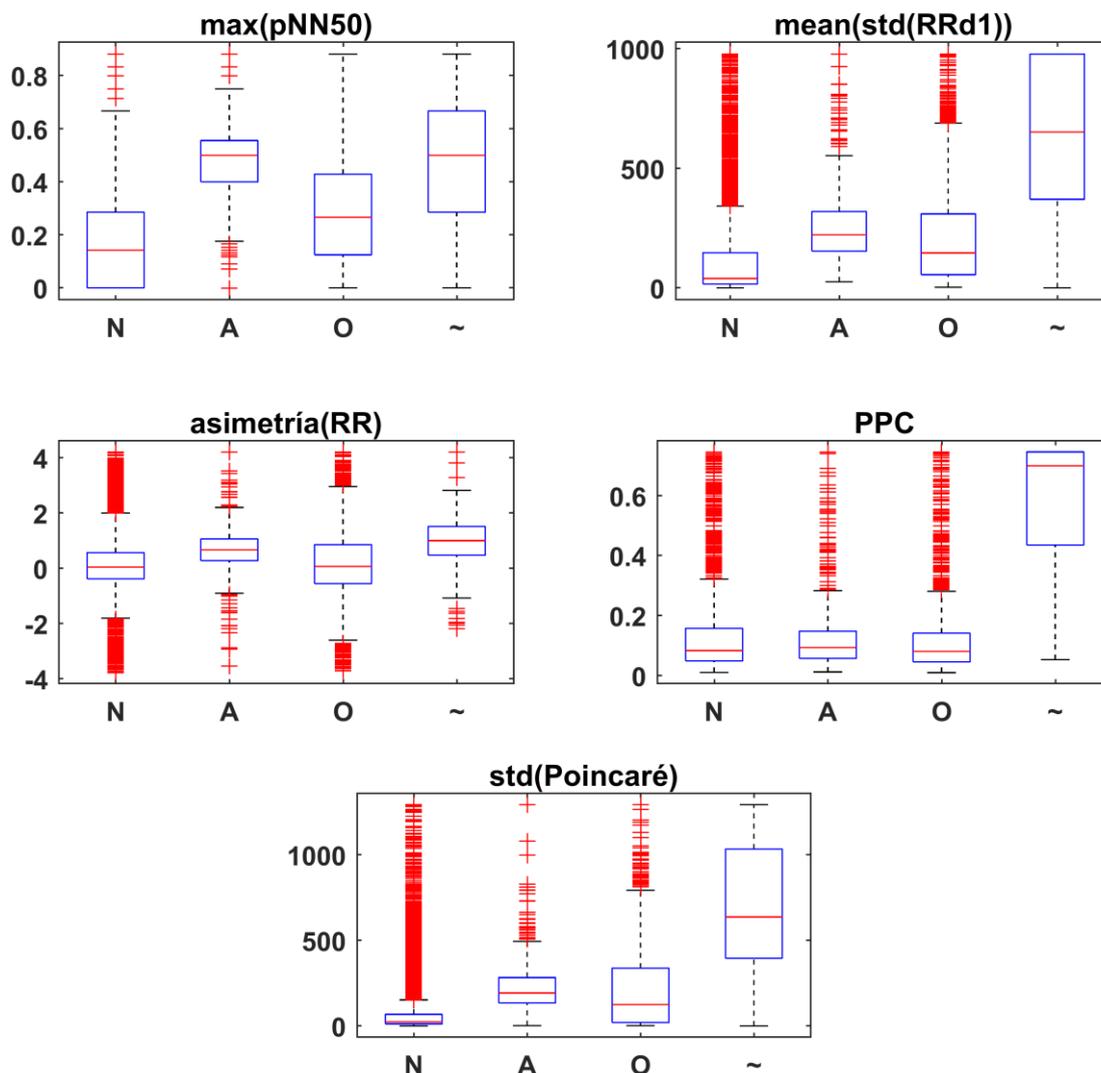


Figura 27. Diagramas de cajas de las cinco variables que sobrevivieron a ambos procesos de selección de características (*Backward Elimination* y *Forward Selection*).

Las trece características restantes seleccionadas mediante *Backward Selection* contenían la medida de distancia  $dtwNNp$  presentada anteriormente. Respecto a las medidas obtenidas utilizando ventana deslizante en la señal, se seleccionó el valor mínimo de la desviación de RRd2; los valores máximos de la desviación de RRd1, RMSD,  $pNN25$  y  $pNN75$ ; los valores medios de la curtosis y media del RR; valores medios del  $pNN75$ ; desviación en la media de la secuencia RR y en la desviación de RRd1. Finalmente, *Backward Elimination* mantuvo como variables a considerar (extraídas utilizando el registro ECG completo, sin ventana deslizante) la media de la secuencia RRd1 y el valor de  $pNN50$ .

En cuanto a las diez características restantes que fueron seleccionadas mediante *Forward Selection*, nos encontramos la medida de distancia  $ecNNo$  entre las variables escogidas.

Respecto a las medidas obtenidas utilizando ventana deslizante en la señal, se seleccionó el valor mínimo de la curtosis de RRd1 y RMSSD del gráfico de Poincaré; los valores máximos de la media del RR; los valores medios de  $pNN25$ ; la desviación en la variable RMSSD. Finalmente, *Forward Selection* mantuvo como variables a considerar (extraídas utilizando el registro ECG completo, sin ventana deslizante) la media de la secuencia RR, la curtosis de RRd1, y la desviación y distancia mínima entre los puntos del gráfico de Lorenz.

A partir de los dos conjuntos de datos anteriores se entrenaron y validaron los distintos modelos de clasificación cuyos resultados presentamos a continuación.

## 7.2. Redes neuronales

Todos los resultados presentados en esta sección fueron obtenidos utilizando una estrategia de validación cruzada de diez particiones.

La Tabla 8 muestra los resultados de las diez redes neuronales que mayor valor de  $F_1$  obtuvieron utilizando el conjunto de 18 variables seleccionadas mediante *Backward Elimination*. El valor  $F_1$  más elevado fue de 0.749, logrado con una red neuronal de una capa oculta compuesta por 512 unidades, un *learning rate* ( $\alpha$ ) de 0.5 y un *momentum* ( $\beta$ ) de 0.4. Se observa entre las diez mejores redes obtenidas aquí que el número de unidades siempre se encuentra entre las 512 y las 1024; de entre el rango de valores explorados, estos se encuentran entre los que requieren más carga computacional, tanto para el entrenamiento como para la clasificación de nuevas muestras por parte del modelo.

**Tabla 8. Mejores resultados obtenidos en la búsqueda de hiperparámetros de las redes neuronales, utilizando validación cruzada de diez particiones y las 18 características seleccionadas mediante *Backward Elimination*. Se muestra el número de unidades de la capa oculta de la red (#Unidades), el *learning rate* ( $\alpha$ ), el *momentum* ( $\beta$ ), y la medida global de *F-Score* ( $F_1$ ), además del rendimiento en cada clase ( $F_{1a}$ ,  $F_{1n}$ ,  $F_{1o}$ ,  $F_{1p}$ ).**

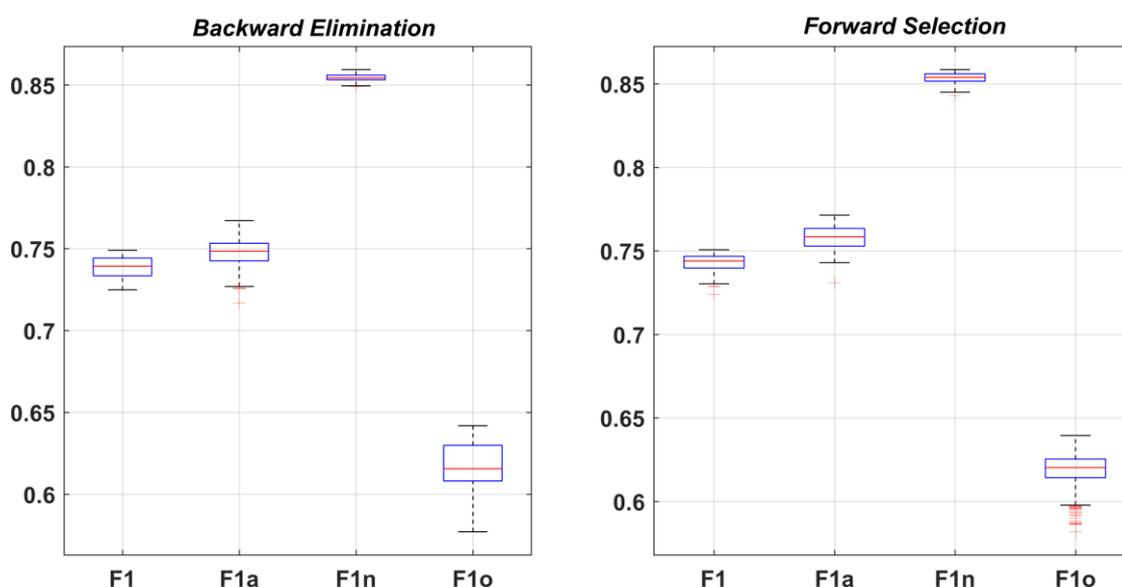
#Unidades	$\alpha$	$\beta$	$F_1$	$F_{1a}$	$F_{1n}$	$F_{1o}$	$F_{1p}$
512	0.50	0.40	<b>0.7490</b>	0.7497	0.8587	0.6386	0.4344
512	0.60	0.00	<b>0.7489</b>	0.7478	0.8572	0.6418	0.4391
512	0.80	0.10	<b>0.7489</b>	0.7489	0.8566	0.6413	0.4141
512	0.50	0.10	<b>0.7487</b>	0.7486	0.8585	0.6389	0.4373
1024	0.50	0.00	<b>0.7484</b>	0.7558	0.8556	0.6338	0.3891
512	0.50	0.00	<b>0.7483</b>	0.7465	0.8583	0.6401	0.4340
1024	0.60	0.10	<b>0.7481</b>	0.7586	0.8532	0.6325	0.3716
1024	0.50	0.20	<b>0.7481</b>	0.7591	0.8536	0.6315	0.3881
1024	0.60	0.20	<b>0.7478</b>	0.7585	0.8527	0.6322	0.3708
1024	0.50	0.30	<b>0.7477</b>	0.7519	0.8551	0.6360	0.3918

La Tabla 9 muestra los resultados de las diez redes neuronales que mayor valor de  $F_1$  obtuvieron utilizando el conjunto de 15 variables seleccionadas mediante *Forward Selection*. El valor  $F_1$  más elevado fue de 0.7506, logrado con 256 unidades en la capa oculta, un *learning rate* ( $\alpha$ ) de 0.6 y un *momentum* ( $\beta$ ) de 0.5. Este modelo, además de mejorar sensiblemente los resultados anteriores, utiliza menos características y número de unidades que el mejor obtenido en el conjunto de variables *Backward Elimination*.

**Tabla 9. Mejores resultados obtenidos en la búsqueda de hiperparámetros de las redes neuronales, utilizando validación cruzada de diez particiones y las 15 características seleccionadas mediante *Forward Selection*. Se muestra el número de unidades de la capa oculta de la red (#Unidades), el *learning rate* ( $\alpha$ ), el *momentum* ( $\beta$ ), y la medida global de *F-Score* ( $F_1$ ), además del rendimiento en cada clase ( $F_{1a}$ ,  $F_{1n}$ ,  $F_{1o}$ ,  $F_{1p}$ ).**

#Unidades	$\alpha$	$\beta$	$F_1$	$F_{1a}$	$F_{1n}$	$F_{1o}$	$F_{1p}$
256	0.60	0.50	<b>0.7506</b>	0.7586	0.8538	0.6395	0.3039
512	0.70	0.20	<b>0.7503</b>	0.7646	0.8554	0.6309	0.4266
256	0.50	0.40	<b>0.7499</b>	0.7573	0.8572	0.6351	0.3734
512	0.70	0.10	<b>0.7498</b>	0.7643	0.8564	0.6286	0.4457
256	0.50	0.30	<b>0.7498</b>	0.7581	0.8566	0.6346	0.3555
256	0.50	0.10	<b>0.7494</b>	0.7623	0.8575	0.6285	0.3597
512	0.80	0.00	<b>0.7494</b>	0.7632	0.8562	0.6288	0.4176
256	1.00	0.50	<b>0.7494</b>	0.7683	0.8517	0.6282	0.2671
256	0.80	0.50	<b>0.7493</b>	0.7622	0.8521	0.6334	0.2985
128	0.60	0.40	<b>0.7493</b>	0.7688	0.8568	0.6222	0.3623

La Figura 28 muestra las medidas *F-Score* de la experimentación en las 180 combinaciones de hiperparámetros de redes neuronales explorados.



**Figura 28. Diagramas de cajas de los rendimientos ( $F_1$ ) de la validación cruzada conseguidos durante la experimentación con redes neuronales. Además aparecen los rendimientos individuales en las clases A, N y O ( $F_{1a}$ ,  $F_{1n}$ ,  $F_{1o}$ ). Se muestran los resultados divididos entre aquellos alcanzados con el conjunto de datos seleccionado mediante *Backward Selection* (18 variables) y *Forward Selection* (15 variables).**

Por otro lado, la Figura 29 presenta los resultados anteriores pero agrupados de manera que se permite comparar en la misma escala cada criterio de evaluación global e individual obtenidos mediante los dos conjuntos de datos seleccionados.

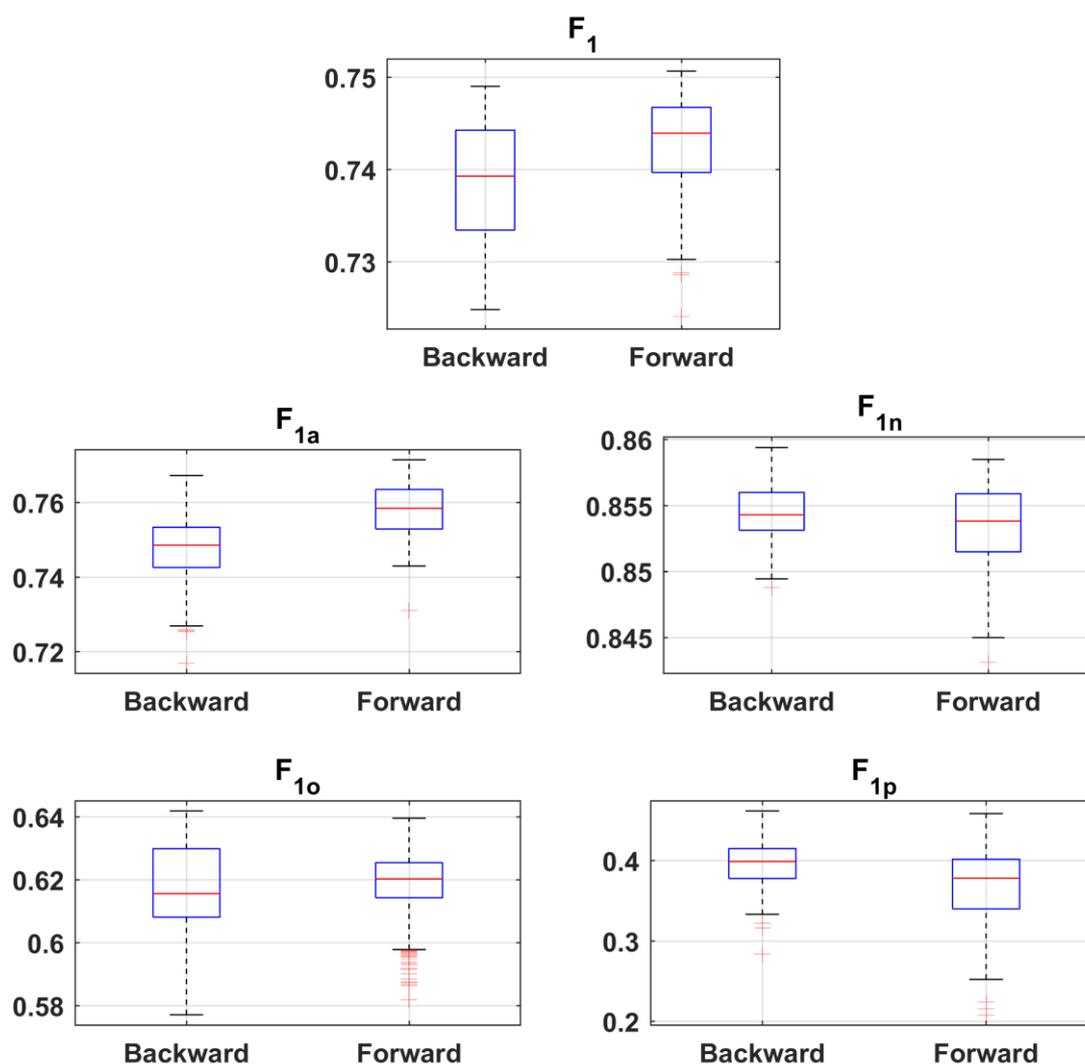


Figura 29. Diagramas de cajas de los rendimientos obtenidos durante la experimentación con redes neuronales agrupados por  $F$ -Score y por conjunto de variables utilizado.

En las figuras anteriores puede apreciarse la existencia de un cierto solapamiento entre los resultados obtenidos en los dos conjuntos de datos. No obstante, se aprecia una ligera mejora en los modelos obtenidos con las 15 variables del *Forward Selection* en la medida de rendimiento global  $F_1$  y en la medida que indica la sensibilidad en la detección de la fibrilación auricular ( $F_{1a}$ ). También se aprecia como el grupo con mayor sensibilidad de detección es el de ritmos normales (N), rondando valores  $F_{1n}$  de 0.85, seguido de cerca por el de fibrilación auricular (A) con valores  $F_{1a}$  cercanos a 0.75. Por otro lado, la detección de otros ritmos anómalos (O) es considerablemente inferior a las anteriores, con valores de  $F_{1o}$  cercanos a 0.62; esto provoca que el rendimiento global del clasificador descienda. Finalmente, la detección de registros ruidosos (~) también se hace complicada, con valores de  $F_{1p}$  cercanos a 0.4; no obstante, esta medida no afecta demasiado al rendimiento global del clasificador, ya que no se encuentra contemplada en el promedio  $F_1$ .

### 7.3. Naïve Bayes

Al igual que en la sección anterior, los resultados presentados aquí fueron obtenidos utilizando una estrategia de validación cruzada de diez particiones.

Tanto en los modelos entrenados con el conjunto de datos obtenido mediante *Backward Elimination*, como en los entrenados con las características de *Forward Selection*, los diez mejores resultados se consiguieron utilizando como probabilidad *a priori* de las clases la que aparecía en el conjunto de entrenamiento (empírica). Además, estos diez mejores resultados también fueron alcanzados usando siempre una función de densidad basada en Kernel (*KDF*) para modelar las distribuciones de probabilidad de cada una de las variables independientes.

La Tabla 10 muestra los resultados de los diez modelos Naïve Bayes que mayor valor de  $F_1$  obtuvieron utilizando el conjunto de 18 variables seleccionadas mediante *Backward Elimination*. El valor  $F_1$  más elevado fue de 0.6595, logrado con un *bandwidth* ( $W$ ) de 0.005 en cada una de las clases. Se observa que no existe un patrón definido que relacione el rendimiento de clasificación con los parámetros  $W$  del modelo.

**Tabla 10. Mejores resultados obtenidos en la experimentación con Naïve Bayes, utilizando validación cruzada de diez particiones y las 18 características seleccionadas mediante *Backward Elimination*. Todos los modelos utilizaron probabilidades de la clase *a priori* calculadas empíricamente y funciones de densidad basadas en Kernel (*KDF*). Se muestran los *bandwidth* empleados para cada una de las clases ( $W_a, W_n, W_o, W_p$ ), así como la medida global de *F-Score* ( $F_1$ ) y el rendimiento en cada clase ( $F_{1a}, F_{1n}, F_{1o}, F_{1p}$ ).**

$[W_a, W_n, W_o, W_p]$	$F_1$	$F_{1a}$	$F_{1n}$	$F_{1o}$	$F_{1p}$
[0.005, 0.005, 0.005, 0.005]	<b>0.6595</b>	0.6577	0.7753	0.5454	0.2583
[0.035, 0.005, 0.005, 0.035]	<b>0.6575</b>	0.6652	0.7736	0.5336	0.3552
[0.050, 0.005, 0.005, 0.050]	<b>0.6574</b>	0.6644	0.7733	0.5345	0.3485
[0.035, 0.005, 0.005, 0.050]	<b>0.6572</b>	0.6651	0.7733	0.5332	0.3530
[0.050, 0.005, 0.005, 0.035]	<b>0.6572</b>	0.6633	0.7736	0.5348	0.3533
[0.005, 0.005, 0.005, 0.020]	<b>0.6565</b>	0.6581	0.7733	0.5381	0.3499
[0.020, 0.005, 0.005, 0.035]	<b>0.6565</b>	0.6663	0.7735	0.5296	0.3592
[0.020, 0.005, 0.005, 0.050]	<b>0.6565</b>	0.6669	0.7733	0.5291	0.3521
[0.005, 0.005, 0.005, 0.035]	<b>0.6562</b>	0.6590	0.7732	0.5363	0.3546
[0.005, 0.005, 0.005, 0.050]	<b>0.6560</b>	0.6584	0.7732	0.5364	0.3464

Por otro lado, la Tabla 11 muestra los resultados de los diez modelos Naïve Bayes que mayor valor de  $F_1$  obtuvieron utilizando el conjunto de 15 variables seleccionadas mediante *Forward Selection*. El valor  $F_1$  más elevado fue de 0.6717, logrado con valores de *bandwidth* ( $W$ ) de 0.05 para fibrilación auricular ( $W_a$ ) y 0.02 para el resto de clases. Aquí también se observa que no existe un patrón definido que relacione el rendimiento de clasificación con los parámetros  $W$  del modelo.

Tabla 11. Mejores resultados obtenidos en la experimentación con Naïve Bayes, utilizando validación cruzada de diez particiones y las 15 características seleccionadas mediante *Forward Selection*. Todos los modelos utilizaron probabilidades de la clase *a priori* calculadas empíricamente y funciones de densidad basadas en Kernel (KDF). Se muestran los *bandwidth* empleados para cada una de las clases ( $W_a, W_n, W_o, W_p$ ), así como la medida global de *F-Score* ( $F_1$ ) y el rendimiento en cada clase ( $F_{1a}, F_{1n}, F_{1o}, F_{1p}$ ).

$[W_a, W_n, W_o, W_p]$	$F_1$	$F_{1a}$	$F_{1n}$	$F_{1o}$	$F_{1p}$
[0.050, 0.020, 0.020, 0.020]	<b>0.6717</b>	0.6855	0.7900	0.5396	0.3373
[0.050, 0.020, 0.020, 0.035]	<b>0.6717</b>	0.6863	0.7897	0.5391	0.3661
[0.050, 0.020, 0.020, 0.050]	<b>0.6717</b>	0.6868	0.7896	0.5385	0.3777
[0.050, 0.005, 0.005, 0.020]	<b>0.6708</b>	0.6858	0.7923	0.5344	0.3515
[0.050, 0.005, 0.005, 0.050]	<b>0.6702</b>	0.6895	0.7914	0.5299	0.3556
[0.035, 0.005, 0.005, 0.020]	<b>0.6701</b>	0.6835	0.7924	0.5343	0.3493
[0.050, 0.005, 0.005, 0.035]	<b>0.6701</b>	0.6884	0.7917	0.5304	0.3590
[0.035, 0.020, 0.020, 0.020]	<b>0.6700</b>	0.6828	0.7898	0.5375	0.3358
[0.050, 0.035, 0.020, 0.020]	<b>0.6698</b>	0.6863	0.7791	0.5441	0.3368
[0.035, 0.020, 0.020, 0.035]	<b>0.6697</b>	0.6828	0.7893	0.5370	0.3642

Finalmente, la Figura 30 muestra las medidas *F-Score* de la experimentación en las 516 combinaciones de configuraciones del modelo Naïve Bayes exploradas. Al igual que en las redes neuronales, existe una mejora en el rendimiento del clasificador cuando se trata de utilizar el conjunto de datos obtenido mediante *Forward Selection*. No obstante, aunque existe una elevada tasa de aciertos entre los ritmos normales (N), en el resto de clases el error de clasificación es elevado, provocando que el rendimiento general de los modelos caiga a valores de  $F_1$  inferiores a 0.7.

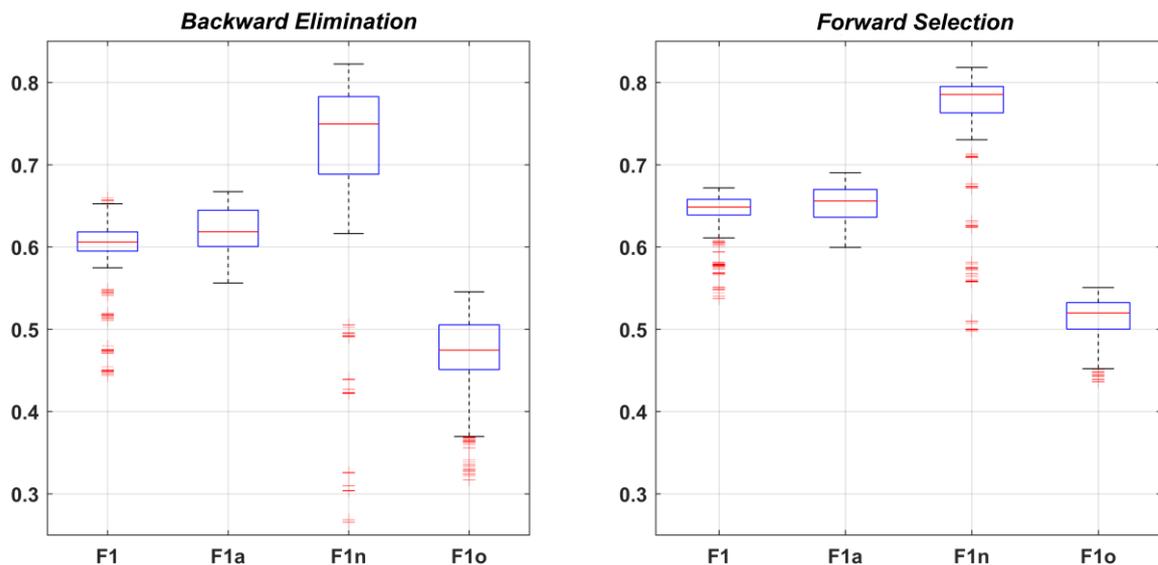
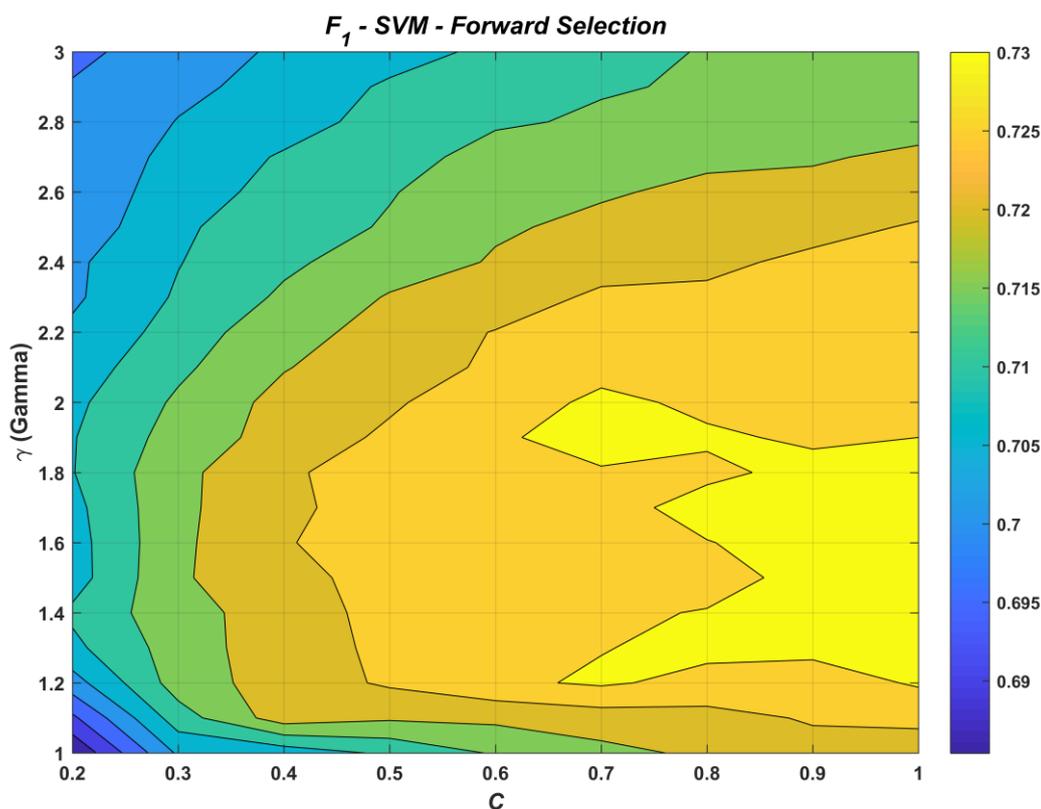


Figura 30. Diagramas de cajas de los rendimientos ( $F_1$ ) de la validación cruzada conseguidos durante la experimentación con Naïve Bayes. Además, aparecen los rendimientos individuales en las clases A, N y O ( $F_{1a}, F_{1n}, F_{1o}$ ). Se muestran los resultados divididos entre aquellos alcanzados con el conjunto de datos seleccionado mediante *Backward Selection* (18 variables) y *Forward Selection* (15 variables).

## 7.4. SVM

Los resultados presentados a continuación fueron alcanzados utilizando una estrategia de validación cruzada de diez particiones, al igual que en los dos tipos de modelos anteriores.

Tanto en los modelos entrenados con el conjunto de datos obtenido mediante *Backward Elimination* como en los entrenados con las características de *Forward Selection* se encontró un patrón en la medida del rendimiento de clasificación  $F_1$  al realizar la búsqueda en la red de parámetros  $C$  y  $\gamma$ . La Figura 31 muestra la tendencia del rendimiento  $F_1$  en una escala de color respecto a los hiperparámetros explorados, encontrándose los valores más elevados en la parte inferior derecha.



**Figura 31.** Patrón de rendimiento de clasificación en la medida  $F_1$  (escala de colores) respecto al espacio de búsqueda en los hiperparámetros de SVM con los datos seleccionados mediante *Forward Selection*. En *Backward Selection* apareció un patrón similar.

Es por ello que, una vez finalizada la primera experimentación, se realizó una segunda búsqueda en el cuadrante de valores  $F_1$  más alto, exactamente en los valores 1.0, 1.1 y 1.2 de  $C$ , y en el rango [1.3, 1.5] de  $\gamma$  en incrementos sucesivos de 0.02. De este modo conseguimos un nuevo grupo de 33 medidas de rendimiento entre los que se encontraba el valor más elevado de  $F_1$ , presentado a continuación en las tablas de resultados. No se pudo ampliar la búsqueda en una zona más extensa debido al coste computacional elevado (cada modelo en esta experimentación realmente requiere del entrenamiento de 6 clasificadores SVM binarios) y la falta de tiempo para ello. No obstante, queda pendiente como trabajo futuro a realizar.

La Tabla 12 muestra los resultados de los diez modelos basados en clasificadores SVM binarios que mayor valor de  $F_1$  obtuvieron utilizando el conjunto de 18 variables seleccionadas mediante *Backward Elimination*. El valor  $F_1$  más elevado fue de 0.7251, logrado con un valor de  $C$  de 1.2 y una función *kernel* de base radial gaussiana con un valor de  $\gamma$  de 1.36.

Por otro lado, la Tabla 13 muestra los resultados de los diez modelos basados en clasificadores SVM binarios que mayor valor de  $F_1$  obtuvieron utilizando el conjunto de 15 variables seleccionadas mediante *Forward Selection*. El valor  $F_1$  más elevado fue de 0.7333, alcanzado con un valor de  $C$  de 1.0 y una función *kernel* de base radial gaussiana con un valor de  $\gamma$  de 1.44.

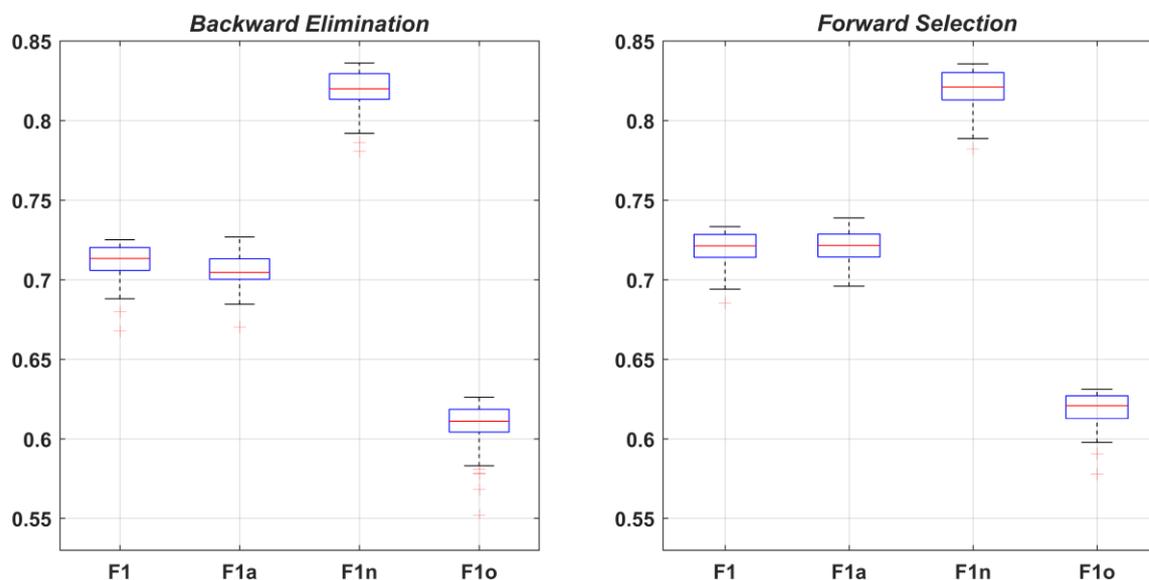
**Tabla 12. Mejores resultados obtenidos en la experimentación con SVM, utilizando validación cruzada de diez particiones y las 18 características seleccionadas mediante *Backward Elimination*. Los kernels utilizados fueron funciones gaussianas de base radial (rbf). Se muestran los valores de *soft margin* ( $C$ ) y factor de escala en la función *rbf* ( $\gamma$ ) empleados, así como la medida global de *F-Score* ( $F_1$ ) y el rendimiento en cada clase ( $F_{1a}$ ,  $F_{1n}$ ,  $F_{1o}$ ,  $F_{1p}$ ).**

$C$	$\gamma$	$F_1$	$F_{1a}$	$F_{1n}$	$F_{1o}$	$F_{1p}$
1.2	1.36	<b>0.7251</b>	0.7262	0.8348	0.6144	0.3371
0.7	1.3	<b>0.7249</b>	0.7217	0.8329	0.6200	0.3716
1.0	1.5	<b>0.7247</b>	0.7211	0.8352	0.6179	0.3858
0.6	1.3	<b>0.7246</b>	0.7193	0.8305	0.6241	0.4027
0.7	1.4	<b>0.7243</b>	0.7210	0.8313	0.6207	0.4235
0.5	1.2	<b>0.7240</b>	0.7184	0.8274	0.6261	0.4108
0.8	1.4	<b>0.7239</b>	0.7216	0.8326	0.6175	0.3903
1.0	1.6	<b>0.7238</b>	0.7197	0.8339	0.6179	0.4203
0.6	1.4	<b>0.7236</b>	0.7172	0.8312	0.6223	0.4747
0.7	1.5	<b>0.7235</b>	0.7168	0.8334	0.6202	0.4885

**Tabla 13. Mejores resultados obtenidos en la experimentación con SVM, utilizando validación cruzada de diez particiones y las 18 características seleccionadas mediante *Forward Selection*. Los kernels utilizados fueron funciones gaussianas de base radial (rbf). Se muestran los valores de *soft margin* ( $C$ ) y factor de escala en la función *rbf* ( $\gamma$ ) empleados, así como la medida global de *F-Score* ( $F_1$ ) y el rendimiento en cada clase ( $F_{1a}$ ,  $F_{1n}$ ,  $F_{1o}$ ,  $F_{1p}$ ).**

$C$	$\gamma$	$F_1$	$F_{1a}$	$F_{1n}$	$F_{1o}$	$F_{1p}$
1.0	1.44	<b>0.7333</b>	0.7358	0.8351	0.6291	0.4354
1.0	1.3	<b>0.7324</b>	0.7388	0.8349	0.6236	0.4052
1.0	1.4	<b>0.7323</b>	0.7343	0.8356	0.6269	0.4250
0.9	1.7	<b>0.7318</b>	0.7308	0.8350	0.6297	0.5051
1.0	1.7	<b>0.7315</b>	0.7315	0.8341	0.6289	0.4818
0.9	1.3	<b>0.7314</b>	0.7377	0.8339	0.6225	0.4147
0.8	1.3	<b>0.7313</b>	0.7364	0.8328	0.6246	0.4204
0.9	1.4	<b>0.7313</b>	0.7312	0.8342	0.6285	0.4323
0.8	1.7	<b>0.7311</b>	0.7297	0.8341	0.6296	0.5036
0.9	1.6	<b>0.7311</b>	0.7311	0.8330	0.6292	0.4627

Asimismo, la Figura 32 muestra las medidas  $F$ -Score de la experimentación en todas las combinaciones de hiperparámetros de SVM explorados, divididos en los dos grupos de datos seleccionados mediante las estrategias *Backward Elimination* y *Forward Selection*.



**Figura 32.** Diagramas de cajas de los rendimientos ( $F_1$ ) de la validación cruzada conseguidos durante la experimentación con SVM. Además, aparecen los rendimientos individuales en las clases A, N y O ( $F_{1a}$ ,  $F_{1n}$ ,  $F_{1o}$ ). Se muestran los resultados divididos entre aquellos alcanzados con el conjunto de datos seleccionado mediante *Backward Selection* (18 variables) y *Forward Selection* (15 variables).

Por otro lado, la Figura 33 presenta estos mismos resultados pero agrupados de manera que se permite comparar en la misma escala cada criterio de evaluación global e individual obtenidos mediante los dos conjuntos de datos seleccionados.

En estas figuras puede apreciarse la existencia de un cierto solapamiento entre los resultados obtenidos en los dos conjuntos de datos. No obstante, al igual que en redes neuronales y Naïve Bayes, se aprecia una mejora en los modelos obtenidos con las 15 variables del *Forward Selection* en la medida de rendimiento global  $F_1$  y en la medida que indica la sensibilidad en la detección de la fibrilación auricular ( $F_{1a}$ ). Además, el rendimiento de clasificación de otros ritmos (O) es también superior utilizando estas mismas variables.

También se aprecia como el grupo con mayor sensibilidad de detección es el de ritmos normales (N), rondando valores  $F_{1n}$  de 0.83, seguido de cerca por el de fibrilación auricular (A), con valores  $F_{1a}$  cercanos a 0.73, algo inferiores que los obtenidos mediante redes neuronales.

Por otro lado, la detección de otros ritmos anómalos (O) es considerablemente inferior a las anteriores, con valores de  $F_{1o}$  cercanos a 0.62; esto provoca que el rendimiento global del clasificador descienda.

Finalmente, la detección de registros ruidosos (–) también se hace complicada, con valores de  $F_{1p}$  en torno a 0.43, no afectando demasiado esta medida al rendimiento global del

clasificador ya que no se encuentra contemplada en el promedio  $F_1$ . Es aquí donde sí que se mejoran ligeramente los rendimientos de clasificación de las redes neuronales.

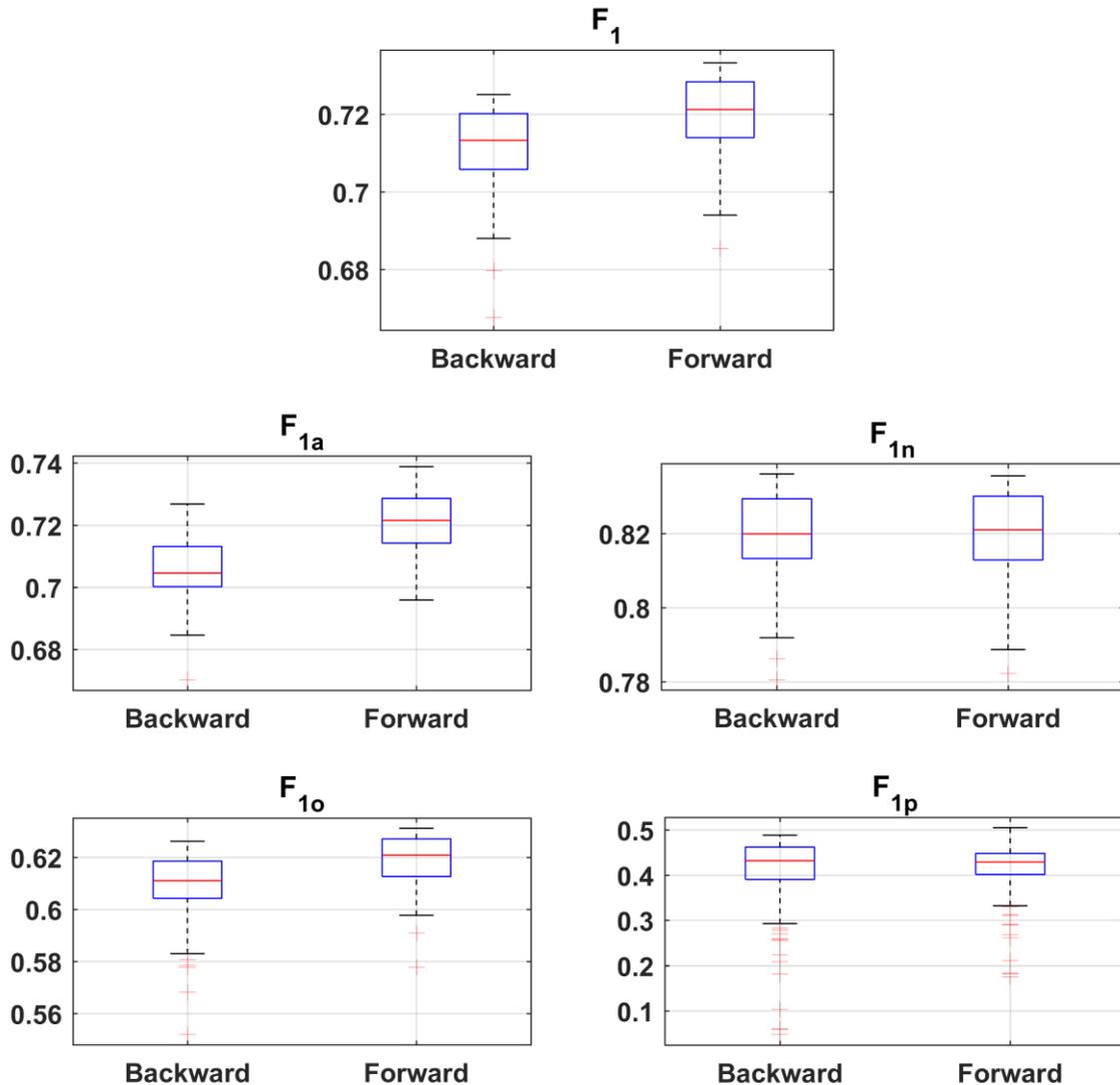


Figura 33. Diagramas de cajas de los rendimientos obtenidos durante la experimentación con SVM agrupados por  $F$ -Score y por conjunto de variables utilizado.

## 7.5. Visión global de los resultados

Una vez presentados todos los resultados obtenidos durante la experimentación con los distintos modelos de clasificación utilizados, en esta sección se presentará una breve visión global del rendimiento de estos clasificadores en nuestros conjuntos de datos.

En cuanto a la selección de características, todos los modelos mostraron siempre un valor  $F_1$  más elevado utilizando las 15 variables seleccionadas mediante *Forward Selection*. Es

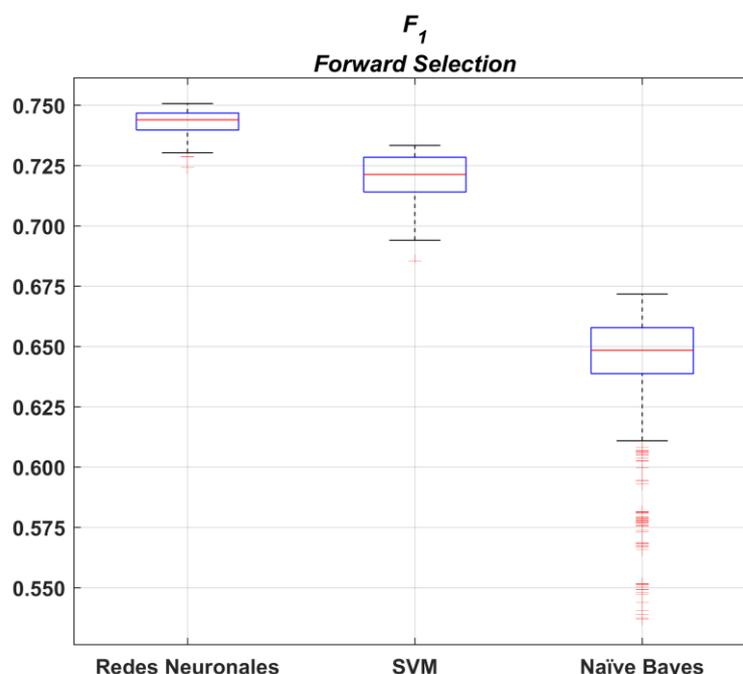
por ello que los resultados presentados a continuación tendrán en cuenta sólo este conjunto de datos.

Respecto al mejor rendimiento de clasificación en cada tipo de modelo, el primer puesto del ranking utilizando la medida  $F_1$  sería para las redes neuronales (0.75), seguidas de cerca por SVM (0.73) y algo más alejado por Naïve Bayes (0.67). Esta norma se cumple también para las puntuaciones individuales en los ritmos normales ( $F_{1n}$ ), de fibrilación auricular ( $F_{1a}$ ) y otros ritmos anómalos ( $F_{1o}$ ). No obstante, en las medidas de sensibilidad de detección de registros ruidosos ( $F_{1p}$ ), SVM mejora considerablemente los resultados de las redes neuronales (0.43 *versus* 0.30); Naïve Bayes también mejora esos resultados, aunque no con tanta diferencia (0.34 *versus* 0.30). La Tabla 14 muestra en detalle el ranking y las medidas  $F$ -Score de los mejores modelos descritos aquí.

**Tabla 14. Ranking de los mejores resultados de clasificación (valor  $F_1$ ) alcanzados durante este trabajo en cada uno de los modelos empleado, así como el rendimiento individual en cada clase ( $F_{1a}$ ,  $F_{1n}$ ,  $F_{1o}$ ,  $F_{1p}$ ).**

Ranking	Modelo	$F_1$	$F_{1a}$	$F_{1n}$	$F_{1o}$	$F_{1p}$
1	Redes Neuronales	<b>0.7506</b>	0.7586	0.8538	0.6395	0.3039
2	SVM	<b>0.7333</b>	0.7358	0.8351	0.6291	0.4354
3	Naïve Bayes	<b>0.6717</b>	0.6855	0.7900	0.5396	0.3373

La Figura 34 muestra el diagrama de cajas de todos los valores  $F_1$  alcanzados con el conjunto de datos *Forward Selection*. Se puede apreciar un ligero solapamiento entre las redes neuronales y SVM, alcanzando un mejor rendimiento en el primer clasificador. Por otro lado, Naïve Bayes queda muy por debajo de los dos anteriores clasificadores, no existiendo solapamiento alguno en sus valores de  $F_1$ .



**Figura 34. Diagrama de cajas de las medidas  $F_1$  obtenidas en todas las experimentaciones de validación cruzada utilizando las 15 variables de *Forward Selection*, agrupadas según el tipo de clasificador.**

## 8. Discusión

---

El principal objetivo de este trabajo fue el entrenamiento y validación de distintos modelos para la clasificación de registros ECG de corta duración y una única derivación en cuatro grupos de ritmos cardiacos distintos: ritmo normal (N), fibrilación auricular (A), otros ritmos (O) y ruidoso (~). Para ello se utilizaron diferentes estrategias para la extracción y selección de características de las señales ECG disponibles, y el posterior entrenamiento y validación de tres tipos distintos de modelos: redes neuronales, SVM y Naïve Bayes.

Los resultados obtenidos durante la extracción de características mostraron diferencias en aquellas obtenidas a partir de la actividad ventricular simple (basadas en la variabilidad del intervalo RR) para diferentes ritmos, incluso cuando la arritmia tenía un origen auricular (FA). Esto permitió evitar problemas asociados con la ausencia de actividad auricular o su enmascaramiento en el ECG, y por lo tanto poder discriminar también la fibrilación auricular del resto de ritmos, sobre todo de los registros con ritmo normal.

Por otro lado, los resultados iniciales obtenidos durante la selección de características mostraron lo que sería la tendencia general durante todo el trabajo: el rendimiento de los clasificadores entrenados con el conjunto de 15 variables seleccionado mediante la estrategia *Forward Selection* siempre resultó mayor que aquel conseguido con las 18 variables seleccionadas mediante *Backward Elimination*. Esta mejora se sitúa en un incremento en torno al 0.01 del valor de  $F_1$ , mejora considerable teniendo en cuenta que además se reduce la dimensionalidad del conjunto de datos a utilizar en tres características. Asimismo, en todos los diagramas de cajas de las medidas *F-Score* correspondientes a *Forward Selection* se observó una menor dispersión en el rendimiento (excepto en SVM); es decir, las cajas son más estrechas que las correspondientes a *Backward Selection*, lo que indica un comportamiento más estable en los modelos entrenados con el primer conjunto de datos.

Por lo tanto, futuros trabajos que utilicen una estrategia de selección de características con la base de datos utilizada deberían emplear la estrategia *Forward Selection*. No obstante, también sería recomendable explorar otras técnicas de selección de características no voraces, como aquellas basadas en algoritmos genéticos o enfriamiento simulado.

Destaca también que, en ambos grupos de variables seleccionadas, siempre resultó escogida alguna característica basada en las medidas de distancia a los vectores  $pNNx$  medios (*dtwNNp* y *ecNNo*), diseñadas durante este trabajo.

Además de lo anterior, durante este trabajo también se exploraron otras técnicas de reducción de la dimensionalidad del conjunto de datos; en concreto, se utilizó un Análisis de Componentes Principales (PCA) para intentar reducir el número de características. No obstante, los resultados que obtuvimos empíricamente eran muy inferiores a los descritos utilizando una selección de características más clásica, por lo que no se han presentado en este estudio.

Respecto a los distintos modelos de clasificación entrenados, los resultados mostraron que los valores de rendimiento más elevados fueron obtenidos mediante las redes neuronales ( $F_1 = 0.75$ ), seguidas de cerca por los modelos SVM ( $F_1 = 0.73$ ). Naïve Bayes, a pesar de presentar un nivel de sensibilidad elevado para los ritmos normales, no consiguió un rendimiento adecuado en el resto de clases en comparación con los otros tipos de modelos ( $F_1 = 0.67$ ).

Los resultados también mostraron que donde se encuentra la mayor dificultad de este problema es en la identificación de otros ritmos anómalos distintos a la fibrilación auricular, así como de los registros ruidosos. Se podría mejorar la identificación de otros ritmos anómalos incorporando características más específicas para estos casos, como, por ejemplo, características relacionadas con el ruido o con otras cardiopatías. Además, se podrían utilizar técnicas de aprendizaje no supervisado o *clustering* para hacer una mayor sub-agrupación de muestras. Estas estrategias podrían mejorar los resultados junto con los métodos propuestos en este trabajo. De este modo, se incrementaría considerablemente el rendimiento de clasificación global de los modelos descritos.

Por otro lado, las experimentaciones y resultados detallados anteriormente se basan en la participación previa del autor en la competición *2017 PhysioNet/Computing in Cardiology Challenge*, cuyos resultados aparecen descritos en [95] y reproducidos en este documento en el Apéndice A – Publicación. Nótese que el mejor valor de  $F_1$  alcanzado en esa publicación durante la validación cruzada en el conjunto de entrenamiento fue de 0.7644; no obstante, ese valor fue obtenido con un etiquetado previo de la base de datos (V2), y por lo tanto menos preciso que el usado en el presente Trabajo Final de Máster (etiquetado V3). Esta circunstancia, junto con el uso de una experimentación ligeramente diferente, explicaría la diferencia de resultados que tenemos en las redes neuronales.

Asimismo, en el Apéndice B – Ranking *Physionet Computing in Cardiology Challenge 2017*, se listan las puntuaciones de los 67 distintos participantes de la competición, así como el enlace a sus publicaciones en aquellos casos donde los autores detallaban su metodología.

Los cuatro participantes que obtuvieron una mayor puntuación  $F_1$  (0.83) utilizaron clasificadores *Random Forest*, estrategias de *Deep Learning*, clasificadores binarios en cascada y una interpretación abductiva de los registros ECG, respectivamente (ver Apéndice B).

En nuestro caso, la posición en el ranking que llegamos a ocupar fue la 34, estando justo en la mitad de la tabla de puntuaciones. De las puntuaciones descritas, destaca que el valor de  $F_1$  que obtuvimos en el conjunto de test (3.658 registros ECG, no accesibles todavía al público hasta la fecha) fue de 0.77. Esto demuestra que nuestro método y las características utilizadas no sobreajustan el rendimiento de los clasificadores al conjunto de entrenamiento, al menos en cuanto a redes neuronales se refiere.

La Figura 35 muestra los diagramas de cajas con las medidas *F-Score* alcanzadas por todos los participantes durante la competición, donde se puede observar como nuestro rendimiento de clasificación  $F_1$  se sitúa justo en el centro del diagrama.

Además, en este mismo gráfico se puede observar como la tendencia en las medidas de sensibilidad individual de detección de las distintas clases es la misma para todos los

participantes, y también la misma que la observada en este trabajo, ordenada de mayor a menor de la siguiente manera:  $F_{1n}$ ,  $F_{1a}$ ,  $F_{1o}$ ,  $F_{1p}$ . Esto demostraría que nuestra metodología es correcta, y que es posible mejorar la clasificación teniendo en cuenta estos patrones de rendimiento, utilizando por ejemplo un sistema de ponderación de errores durante el entrenamiento de nuevos modelos.

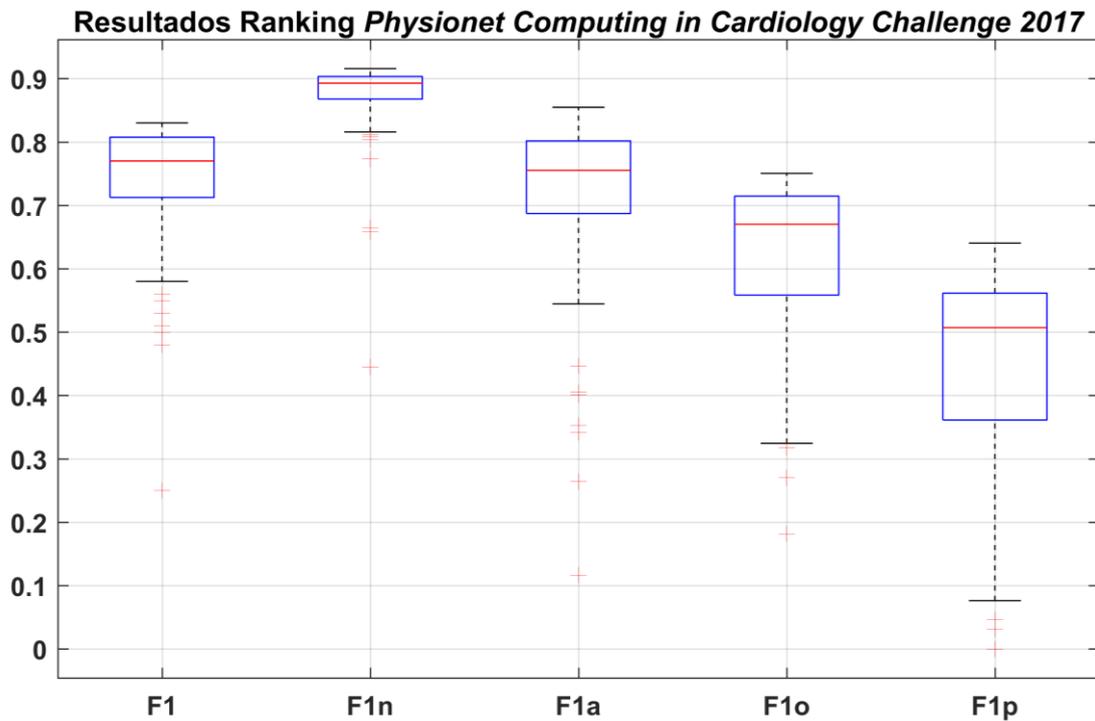


Figura 35. Diagramas de caja con los resultados en las medidas de *F-Score* de los 67 participantes en la competición *2017 PhysioNet/Computing in Cardiology Challenge*, utilizando el conjunto de test compuesto por 3.658 registros ECG no accesibles todavía al público hasta la fecha (ver Apéndice B).



## 9. Conclusiones

---

La incidencia y prevalencia de las arritmias cardíacas, entre las que destaca la fibrilación auricular, está aumentando. Por otro lado, hoy en día los dispositivos portátiles o “*wearables*” están ganando gran interés como dispositivos de monitorización, tanto en investigación como en ámbitos clínicos. No obstante, los métodos automáticos para diagnosticar de manera fiable la FA utilizando las señales ECG proporcionadas por dispositivos portátiles son todavía un reto, especialmente si también se consideran otros ritmos normales o patológicos.

En el presente trabajo, se realizó en primer lugar una revisión bibliográfica de las posibles características a extraer de la señal ECG y de las técnicas utilizadas en otros trabajos para la detección de la fibrilación auricular.

A continuación, presentamos y evaluamos un conjunto de modelos para la clasificación de registros ECG de corta duración y una única derivación en cuatro grupos o clases de ritmos cardíacos distintos: ritmo normal (N), fibrilación auricular (A), otros ritmos (O) y ruidoso (~).

Para ello combinamos la extracción y selección de características ventriculares, basadas en su mayoría en la variabilidad del intervalo RR, con el entrenamiento y validación de tres tipos de modelos de clasificación: redes neuronales, Naïve Bayes y SVM. Todo esto utilizando una metodología robusta de evaluación del rendimiento de clasificación mediante medidas *F-Score* y validación cruzada de diez particiones.

Los resultados mostraron que los modelos que mejor rendimiento presentaban en la clasificación de los distintos ritmos fueron las redes neuronales ( $F_1 = 0.75$ ), seguidas de cerca por los modelos SVM ( $F_1 = 0.73$ ). Naïve Bayes, a pesar de presentar un nivel de sensibilidad elevado para los ritmos normales, no consiguió un rendimiento adecuado en el resto de clases en comparación con los otros tipos de modelos ( $F_1 = 0.67$ ).

Los resultados también mostraron que donde se encuentra la mayor dificultad de este problema es en la identificación de otros ritmos anómalos distintos a la fibrilación auricular, así como de los registros ruidosos. Mejorar la identificación de otros ritmos anómalos incorporando características más específicas para estos casos en el método propuesto sería una de las direcciones interesantes a explorar en trabajos futuros. De este modo, se incrementaría considerablemente el rendimiento de clasificación global de los modelos descritos.

En resumen, los registros ECG utilizados durante este trabajo comparten muchas características con los adquiridas por distintos dispositivos *wearables* en el mercado. Por otro lado, los modelos entrenados y evaluados aquí presentan un bajo consumo computacional durante la clasificación. Es por esto por lo que los clasificadores resultantes podrían ser buenos candidatos para ser implementados en sistemas de gestión de pacientes o Sistemas de Ayuda al Diagnóstico Médico que utilicen dispositivos portátiles para la monitorización cardíaca.



## 10. Trabajos futuros

---

La lista de líneas futuras de trabajo para este estudio es larga y muy interesante, además de prometedora.

En primer lugar, y como ya se ha descrito anteriormente, es necesario incrementar la sensibilidad en la detección de los registros que presentan ritmos anómalos distintos a la fibrilación auricular. Para ello sería necesario extraer de la señal ECG otro tipo de características no relacionadas con la variabilidad del RR, como, por ejemplo, la ausencia de onda P, o su amplitud y morfología. Lo mismo sucede con características basadas en la amplitud y morfología del complejo QRS-T. Un detector de latidos ectópicos también sería útil en este sentido. Finalmente, utilizar características relacionadas con la frecuencia, o el tiempo-frecuencia de la señal también sería posible, aunque en registros de una única derivación podría no ser una estrategia adecuada, además de su elevado coste computacional, inviable para dispositivos móviles. Otro espacio a explorar en este sentido sería el diseño y desarrollo de nuevas características de la señal ECG que aportaran más información concreta a este problema.

En segundo lugar, sería interesante utilizar otras técnicas de selección de características, como selección aleatoria, algoritmos genéticos, enfriamiento simulado u optimización por enjambre de partículas, entre otras. De este modo, sería posible encontrar combinaciones de variables independientes que permitieran un mejor rendimiento de clasificación, y a las que no es posible llegar con los métodos voraces utilizados en este trabajo. En este sentido, también sería interesante realizar una selección de características independiente y *ad-hoc* para cada uno de los modelos de clasificación a utilizar, encontrando la combinación óptima concreta que mejores resultados le proporcione.

En tercer lugar, la ampliación de la base de datos de entrenamiento permitiría hacer que los modelos tuvieran más variabilidad estadística sobre la que poder tomar mejores decisiones. Existen bases de datos con registros ECG ya etiquetados, incluso latido a latido, de las que extraer información muy valiosa al respecto. Muchas de ellas se encuentra indizadas en el *PhysioBank Databases* de *PhysionNet* (<https://physionet.org/physiobank/database/>). No obstante, esto requeriría un fuerte preprocesado de los datos para obtener un conjunto de registros y etiquetas consistente.

Finalmente, y como línea de trabajo futura más interesante, encontramos la experimentación con otros tipos de modelos o técnicas de clasificación. Respecto a los modelos de clasificación que no se emplearon durante este estudio, sería interesante comprobar el rendimiento de las Redes Bayesianas, puesto que también pueden encontrar dependencias probabilísticas [98]. Por otro lado, los modelos ocultos de Markov son ampliamente utilizados en problemas de reconocimiento del habla, bioinformática y otros campos de análisis de secuencias o señales. Es por ello que la extrapolación a la clasificación de estos registros ECG sería muy interesante y podría ayudar a mejorar el rendimiento de los clasificadores clásicos. Lo mismo sucede con las técnicas de aprendizaje profundo o *Deep Learning* actuales.



## 11. Bibliografía

---

- [1] J. Montes-Santiago, V. Rodil, F. Formiga, J. Cepeda y A. Urrutia, «Características y costes de los pacientes ingresados por arritmias cardíacas en España» *Revista Clínica Española*, vol. 213, nº 5, pp. 235-9, 2013.
- [2] G. Chow, J. Marine y J. Fleg, «Epidemiology of Arrhythmias and Conduction Disorders in Older Adults» *Clinics in Geriatric Medicine*, vol. 28, nº 4, pp. 539-53, 2012.
- [3] S. Chugh, R. Havmoeller, K. Narayanan, D. Singh, M. Rienstra, E. Benjamin y C. Murray, «Worldwide Epidemiology of Atrial Fibrillation: A Global Burden of Disease 2010 Study» *Circulation*, vol. 129, nº 8, pp. 837-47, 2014.
- [4] N. Wheeldon, «Atrial fibrillation and anticoagulant therapy» *European Heart Journal*, vol. 16, nº 3, pp. 302-12, 1995.
- [5] G. Naccarelli, H. Varker, J. Lin y K. Schulman, «Increasing prevalence of atrial fibrillation and flutter in the United States» *American Journal of Cardiology*, vol. 104, nº 11, pp. 1534-39, 2009.
- [6] P. Kumari, L. Mahew y P. Syal, «Increasing trend of wearables and multimodal interface for human activity monitoring: A review.» *Biosensors and Bioelectronics*, vol. 90, pp. 298-307, 2017.
- [7] D. Eddy, «Evidence-based medicine: a unified approach,» *Health Aff (Millwood)*, vol. 24, nº 1, pp. 9-17, 2005.
- [8] H. Grain, *Guide to the principles and desirable features of clinical decision support*, Standards Australia, 2007.
- [9] C. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.
- [10] G. Clifford, C. Liu, B. Moody, I. Silva, Q. Li, A. Johnson y R. Mark, «AF Classification from a Short Single Lead ECG Recording: the PhysioNet Computing in Cardiology Challenge 2017» de *Computing in Cardiology*, Rennes (France), 2017.
- [11] A. Guyton y J. Hall, *Textbook of Medical Physiology*, 11 ed, Elsevier Saunders, 2006.
- [12] K. Saladin, *Anatomía y Fisiología. La unidad entre forma y función*, 6a ed, McGraw-Hill, 2012.
- [13] J. Hume y A. Grant, «Agents Used in Cardiac Arrhythmias» de *Basic & Clinical Pharmacology*, 13e, McGraw-Hill Medical, 2012.

- [14] G. Lip, L. Fauchier, S. Freedman, I. Gelder, A. Natale, C. Gianni, S. Nattel, T. Potpara, M. Rienstra, H. Tse y D. Lane, «Atrial fibrillation» *Nature Reviews Disease Primers* 2, vol. 31, nº 2, 2016.
- [15] A. Camm, P. Kirchhof, G. Lip, U. Schotten, I. Savelieva, S. Ernst, I. Gelder, N. Al-Attar, G. Hindricks, B. Prendergast, H. Heidbuchel, O. Alfieri, A. Angelini, D. Atar, P. Colonna, R. Caterina, J. Sutter, A. Goette, B. Gorenek, M. Heldal, S. Hohloser, P. Kolh, J. Heuzey, P. Ponikowski y F. Rutten, «Guidelines for the management of atrial fibrillation: the Task Force for the Management of Atrial Fibrillation of the European Society of Cardiology (ESC)» *European Heart Journal*, vol. 31, nº 19, pp. 2369-429, 2010.
- [16] L. Mainardi, L. Sörnmo y S. Cerutti, *Understanding Atrial Fibrillation: the Signal Processing Contribution*, Morgan and Claypool Publishers, 2008.
- [17] I. Savelieva y J. Camm, «Update on atrial fibrillation: part I» *Clinical Cardiology*, vol. 31, nº 2, pp. 55-62, 2008.
- [18] V. Fuster, L. Ryden, R. Asinger, D. Cannom y H. Crijns, «ACC/AHA/ESC guidelines for the management of patients with atrial fibrillation: executive summary» *Journal of the American College of Cardiology*, vol. 38, nº 4, pp. 1231-65, 2001.
- [19] J. Gordon, K. Young y D. Kruse, *Anatomy & Physiology*, openStax, 2013.
- [20] M. Allesie, K. Konings y M. Wijffels, *Atrial Arrhythmias—State of the Art: Electrophysiological Mechanism of Atrial Fibrillation*, NY: Eds. Armonk, NY: Futura Publ. Co., 1995.
- [21] J. Jalife, «Los rotores y la fibrilación» *Archivos de Cardiología de México*, vol. 74, nº 2, pp. 290-2, 2004.
- [22] G. Moe, «On the multiple wavelet hypothesis of atrial fibrillation» *Archives Internationales De Pharmacodynamie Et De Therapie*, vol. 140, pp. 183-8, 1962.
- [23] J. Jalife, O. Berenfeld, A. Skanes y R. Mandapati, «Mechanisms of atrial fibrillation: mother rotors or multiple daughter wavelets, or both?» *Journal of Cardiovascular Electrophysiology*, vol. 9, nº 8, pp. S2-12, 1998.
- [24] J. Jalife, O. Berenfeld y M. Mansour, «Mother rotors and fibrillatory conduction: a mechanism of atrial fibrillation» *Cardiovascular Research*, vol. 54, nº 2, pp. 204-16, 2002.
- [25] M. Vaquero, D. Calvo y J. Jalife, «Cardiac fibrillation: From ion channels to rotors in the human heart» *Heart Rhythm*, vol. 5, nº 6, pp. 872-9, 2008.
- [26] A. Bollmann, K. Binias, I. Toepffer, J. Molling, C. Geller y H. Klein, «Importance of left atrial diameter and atrial fibrillatory frequency for conversion of persistent atrial

- fibrillation with oral flecainide» *American Journal of Cardiology*, vol. 90, nº 9, pp. 1011-4, 2002.
- [27] A. Bollmann, N. Kanuru, K. McTeague, P. Walter, D. DeLurgio y J. Langberg, «Frequency analysis of human atrial fibrillation using the surface electrocardiogram and its response to ibutilide» *American Journal of Cardiology*, vol. 81, nº 12, pp. 1439-45, 1998.
- [28] L. Gaztañaga, F. Marchlinski y B. Betensky, «Mecanismos de las arritmias cardiacas» *Revista Española de Cardiología*, vol. 65, nº 2, pp. 174-85, 2012.
- [29] S. Ladavich y B. Ghoraani, «Rate-independent detection of atrial fibrillation by statistical modeling of atrial activity» *Biomedical Signal Processing and Control*, vol. 18, nº 4, pp. 274-81, 2015.
- [30] X. Du, N. Rao, M. Qian, D. Liu, J. Li, W. Feng, L. Yin y X. Chen, «A novel method for real-time atrial fibrillation detection in electrocardiograms using multiple parameters» *Annals of Noninvasive Electrocardiology*, vol. 19, nº 3, pp. 217-25, 2014.
- [31] H. Pürerfellner, E. Pokushalov, S. Sarkar, J. Koehler, R. Zhou, L. Urban y G. Hindricks, «P-wave evidence as a method for improving algorithm to detect atrial fibrillation in insertable cardiac monitors» *Heart Rhythm*, vol. 11, nº 9, pp. 1575-83, 2014.
- [32] R. Alcaraz, C. Vaya, R. Cervigon, C. Sanchez y J. Rieta, «Wavelet sample entropy: A new approach to predict termination of atrial fibrillation» *Computing in Cardiology*, pp. 597-600, 2006.
- [33] J. Ródenas, M. García, R. Alcaraz y J. Rieta, «Wavelet entropy automatically detects episodes of atrial fibrillation from single-lead electrocardiograms» *Entropy*, vol. 17, nº 9, pp. 6179-99, 2015.
- [34] M. García, J. Ródenas, R. Alcaraz y J. Rieta, «Application of the relative wavelet energy to heart rate independent detection of atrial fibrillation» *Computer Methods and Programs in Biomedicine*, vol. 131, nº 7, pp. 157-68, 2016.
- [35] A. Petrénas, V. Marozas, L. Sörnmo y A. Lukosevicius, «An echo state neural network for QRST cancellation during atrial fibrillation» *IEEE Transactions on Biomedical Engineering*, vol. 59, nº 10, pp. 2950-57, 2012.
- [36] J. Rieta, F. Castells, C. Sánchez, V. Zarzoso y J. Millet, «Atrial activity extraction for atrial fibrillation analysis using blind source separation» *IEEE Transactions on Biomedical Engineering*, vol. 51, pp. 1176-86, 2004.
- [37] J. Slocum, A. Sahakian y S. Swiryn, «Diagnosis of atrial fibrillation from surface electrocardiograms based on computer-detected atrial activity» *Journal of Electrocardiology*, vol. 25, pp. 1-8, 1992.

- [38] D. Raine, P. Langley, A. Murray, A. Dunuwille y J. Bourke, «Surface atrial frequency analysis in patients with atrial fibrillation» *Journal of Cardiovascular Electrophysiology*, vol. 15, pp. 1021-6, 2004.
- [39] F. Castells, C. Mora, J. Rieta, D. Moratal-Pérez y J. Millet, «Estimation of atrial fibrillatory wave from single-lead atrial fibrillation electrocardiograms using principal component analysis concepts.» *Medical & Biological Engineering & Computing*, vol. 43, nº 5, pp. 557-60, 2005.
- [40] F. Castells, J. Rieta, J. Millet y V. Zarzoso, «Spatiotemporal Blind Source Separation Approach to Atrial Activity Estimation in Atrial Tachyarrhythmias» *IEEE Transactions on Biomedical Engineering*, vol. 52, pp. 258-67, 2005.
- [41] C. Vasquez, A. Hernandez, F. Mora, G. Carrault y G. Passariello, «Atrial activity enhancement by wiener filtering using an artificial neural network» *IEEE Transactions on Biomedical Engineering*, vol. 48, nº 8, pp. 940-4, 2001.
- [42] J. Mateo y J. Rieta, «Radial basis function neural networks applied to efficient QRST cancellation in atrial fibrillation» *Computers in Biology and Medicine*, vol. 43, nº 2, pp. 154-63, 2013.
- [43] E. Ciaccio, A. Biviano, W. Whang y H. Garan, «A new LMS algorithm for analysis of atrial fibrillation signals» *BioMEDICAL Engineering Online*, vol. March, pp. 11-15, 2012.
- [44] A. Bollmann, D. Husser, L. Mainardi, F. Lombardi, P. Langley, A. Murray, J. Rieta, J. Millet, S. Olsson, M. Stridh y L. Sörnmo, «Analysis of surface electrocardiograms in atrial fibrillation: Techniques, research, and clinical applications» *Europace*, vol. 8, nº 11, pp. 911-26, 2006.
- [45] M. Stridh, L. Sörnmo, C. Meurling y S. Olsson, «Characterization of atrial fibrillation using the surface ECG: time-dependent spectral properties» *IEEE Transactions on Biomedical Engineering*, vol. 48, pp. 19-27, 2001.
- [46] F. Sandberg, M. Stridh y L. Sörnmo, «Frequency Tracking of Atrial Fibrillation Using Hidden Markov Models» *IEEE Transactions on Biomedical Engineering*, vol. 55, pp. 502-11, 2008.
- [47] R. Colloca, L. Minardi y G. Clifford, Implementation and testing of atrial fibrillation detectors for a mobile phone application, Politecnico di Milano and University of Oxford, 2013.
- [48] M. Carrara, L. Carozzi, T. Moss, M. Pasquale, S. Cerutti, M. Ferrario, D. Lake y J. Moorman, «Heart rate dynamics distinguish among atrial fibrillation, normal sinus rhythm and sinus rhythm with frequent ectopy» *Physiological Measurement*, vol. 36, nº 9, pp. 1873-88, 2015.

- [49] J. Park, S. Lee y M. Jeon, «Atrial fibrillation detection by heart rate variability in Poincare plot» *BioMedical Engineering OnLine*, vol. 8, nº 38, 2009.
- [50] S. Sarkar, D. Ritscher y R. Mehra, «A detector for a chronic implantable atrial tachyarrhythmia monitor» *IEEE Transactions on Biomedical Engineering*, vol. 55, nº 3, pp. 1219-24, 2008.
- [51] L. Zhang, T. Guo, B. Xi, Y. Fan y Y. Wang, «Automatic recognition of cardiac arrhythmias based on the geometric patterns of Poincaré plots» *Physiological Measurement*, vol. 36, pp. 283-301, 2015.
- [52] K. Tateno y L. Glass, «Automatic detection of atrial fibrillation using the coefficient of variation and density histograms of RR and deltaRR intervals» *Medical & Biological Engineering & Computing*, vol. 39, nº 6, pp. 664-671, 2001.
- [53] D. Linker, «Accurate, Automated Detection of Atrial Fibrillation in Ambulatory Recordings» *Cardiovascular Engineering and Technology*, vol. 7, nº 2, pp. 182-9, 2016.
- [54] C. Huang, S. Ye, H. Chen, D. Li, F. He y Y. Tu, «A novel method for detection of the transition between atrial fibrillation and sinus rhythm» *IEEE Transactions on Biomedical Engineering*, vol. 58, nº 4, pp. 1113-9, 2011.
- [55] S. Cerutti, L. Mainardi y L. Sörnmo, *Understanding atrial fibrillation: the signal processing contribution*, Morgan and Claypool Publishers, 2008.
- [56] A. Petrénas, V. Marozas y L. Sörnmo, «Low-complexity detection of atrial fibrillation in continuous long-term monitoring» *Computers in Biology and Medicine*, vol. 65, nº 10, pp. 184-91, 2015.
- [57] X. Zhou, H. Ding, B. Ung, E. Pickwell-MacPherson y Y. Zhang, «Automatic online detection of atrial fibrillation based on symbolic dynamics and Shannon entropy» *BioMedical Engineering OnLine*, vol. 13, nº 18, 2014.
- [58] R. Alcaraz, D. Abásolo, R. Hornero y J. Rieta, «Optimal parameters study for sample entropy-based atrial fibrillation organization analysis» *Computer Methods and Programs in Biomedicine*, vol. 99, nº 1, pp. 124-32, 2010.
- [59] D. Lake y J. Moorman, «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*, vol. 300, nº 1, pp. 319-25, 2011.
- [60] D. DeMazumder, D. Lake, A. Cheng, T. Moss, E. Guallar, R. Weiss, S. Jones, G. Tomaselli y J. Moorman, «Dynamic analysis of cardiac rhythms for discriminating atrial fibrillation from lethal ventricular arrhythmias» *Circulation: Arrhythmia and Electrophysiology*, vol. 6, nº 3, pp. 555-61, 2013.

- [61] B. Anuradha y V. Veera, «Cardiac Arrhythmia Classification using Fuzzy Classifiers» *Journal of Theoretical and Applied Information Technology*, pp. 353-9, 2008.
- [62] S. Babaeizadeh, R. Gregg, E. Helfenbein, J. Lindauer y S. Zhou, «Improvements in atrial fibrillation detection for real-time monitoring» *Journal of Electrocardiology*, vol. 42, nº 6, pp. 522-6, 2009.
- [63] A. Petrénas, L. Sörnmo, A. Lukoševicius y V. Marozas, «Detection of occult paroxysmal atrial fibrillation» *Medical & Biological Engineering & Computing*, vol. 53, nº 4, pp. 287-97, 2015.
- [64] J. Oster y C. GD, «Impact of the presence of noise on RR interval-based atrial fibrillation detection» *Journal of Electrocardiology*, vol. 48, nº 6, pp. 947-51, 2015.
- [65] R. Colloca, A. Johnson, L. Mainardi y G. Clifford, «A support vector machine approach for reliable detection of atrial fibrillation events» de *Computing in Cardiology*, Zaragoza (Spain), 2013.
- [66] Q. Li, C. Liu, Oster y G. Clifford, «Signal processing and feature selection preprocessing for classification in noisy healthcare data» de *Machine Learning for Healthcare Technologies*, David A Clifton, 2016, pp. 33-58.
- [67] M. Guillem, A. Climent, F. Castells, D. Husser, J. Millet, A. Arya, C. Piorkowski y A. Bollmann, «Noninvasive mapping of human atrial fibrillation» *Journal of Cardiovascular Electrophysiology*, vol. 20, pp. 507-13, 2009.
- [68] M. Guillem, A. Climent, J. Millet, Á. Arenal, F. Fernández-Avilés, J. Jalife, F. Atienza y O. Berenfeld, «Noninvasive localization of maximal frequency sites of atrial fibrillation by Body Surface Potential Mapping» *Circulation: Arrhythmia and Electrophysiology*, vol. 6, pp. 294-301, 2013.
- [69] A. Shah, M. Hocini, M. Haissaguerre y P. Jais, «Non-invasive Mapping of Cardiac Arrhythmias» *Current Cardiology Reports*, vol. 8, pp. 17-60, 2015.
- [70] A. Pullan, M. Buist, L. Cheng y M. Buist, *Mathematically modelling the electrical activity of the heart: from cell to body surface and back again*, Singapore: World Scientific, 2005.
- [71] A. Liberos, *Mathematical modeling approaches for the diagnosis and treatment of reentrant atrial tachyarrhythmias*, Universitat Politècnica de València, 2016.
- [72] L. Martínez-Mateu, *Mapping of the Electrical Activity of Human Atria. Multiscale Modelling and Simulations*, Universitat Politècnica de València, 2018.
- [73] C. Tobon, M. Duarte, J. Duque, M. Becerra, S. Arango, K. Cardona y J. Saiz, «Chloroquine effect on rotor termination under paroxysmal and chronic atrial

- fibrillation. 2D simulation study» de *IEEE Central America and Panama Convention (CONCAPAN XXXIV)*, 2014.
- [74] S. Kharche, T. Stary, M. Colman, I. Biktasheva, A. Workman, A. Rankin, A. Holden y H. Zhang, «Effects of human atrial ionic remodelling by  $\beta$ -blocker therapy on mechanisms of atrial fibrillation: a computer simulation» *Europace*, vol. 16, nº 10, pp. 1524-33, 2014.
- [75] G. Box y D. NR, *Empirical Model-Building and Response Surfaces*, Wiley, 1987, p. 424.
- [76] L. Breiman, J. Friedman, R. Olshen y C. Stone, *Classification and Regression Trees*, Boca Raton: CRC Press, 1984.
- [77] B. Dasarathy, *Nearest Neighbor (NN) Norms: NN Pattern Classification Techniques*, IEEE Computer Society, 1990.
- [78] S. Tortajada, J. García-Gómez y M. Robles-Viejo, *Incremental Learning approaches to Biomedical decision problems*, Valencia: Universitat Politècnica de València, 2012.
- [79] F. Rosenblatt, «The Perceptron: a probabilistic model for information storage and organization in the brain.» *Psychological Review*, vol. 65, nº 6, pp. 386-408, 1958.
- [80] Y. Hu y J. Hwang, *Handbook of neural networks signal processing*, CRC Press, 2001.
- [81] I. Gondra, «Applying machine learning to software fault-proneness prediction» *The Journal of Systems and Software*, vol. 81, pp. 186-95, 2008.
- [82] C. Bishop, *Neural Networks for Pattern Recognition.*, Oxford: Clarendon Press, 1995.
- [83] M. Zeiler, M. Ranzato y R. Monga, «On rectified linear units for speech processing» *Speech and Signals Processing (ICASSP)*, nº 1, pp. 3517-21, 2013.
- [84] C. G, «Approximation by superpositions of a sigmoidal function» *Mathematics of control, signals and systems*, vol. 2, nº 4, pp. 303-14, 1989.
- [85] T. Mitchell, *Machine Learning*, McGraw Hill, 1997.
- [86] J. Shawe-Taylor y N. Cristianini, *An introduction to Support Vector Machines and other kernel-based learning methods*, Cambridge University Press, 2000.
- [87] C. Hsu y L. CJ, «A comparison of methods for multiclass support vector machines» *IEEE transactions on Neural Networks*, vol. 13, nº 2, pp. 415-25, 2002.
- [88] R. Palm, *Prediction as a candidate for learning deep hierarchical models of data*, Technical University of Denmark, 2012.

- [89] N. Japkowicz y S. Stephen, «The class imbalance problem: a systematic study.» *Intelligent data analysis Journal*, vol. 6, nº 5, pp. 429-49, 2002.
- [90] Y. Sasaki, «The truth of the F-measure» University of Manchester, 2007.
- [91] H. Tsuji, M. Larson, F. Venditti, E. Manders, J. Evans, C. Feldman y D. Levy, «Impact of Reduced Heart Rate Variability on Risk for Cardiac Events: The Framingham Heart Study» *Circulation*, vol. 94, pp. 2850-55, 1996.
- [92] F. Shaffer y J. Ginsberg, «An Overview of Heart Rate Variability Metrics and Norms» *Frontiers in Public Health*, vol. 5, nº 258, pp. 1-17, 2017.
- [93] J. Mietus, C. Peng, I. Henry, R. Goldsmith y A. Goldberger, «The pNNx files: re-examining a widely used heart rate variability measure» *Heart*, vol. 88, pp. 378-380, 2002.
- [94] X. Ruan, C. Liu, C. Liu, X. Wang y P. Li, «Automatic Detection of Atrial Fibrillation Using R-R Interval Signal» de *4th International Conference on Biomedical Engineering and Informatics (BMEI)*, 2011.
- [95] S. Jiménez-Serrano, J. Yagüe, E. Simarro, C. Calvo, F. Castells y J. Millet, «Atrial Fibrillation Detection Using Feedforward Neural Networks and Automatically Extracted Signal Features» de *Computing in Cardiology (CinC) 2017*, Rennes (France), 2017.
- [96] D. Pyle, *Data Preparation for Data Mining*, Morgan Kaufmann Publishers, 1999.
- [97] «Matlab R2013a Online Documentation,» [En línea]. Available: <http://www.mathworks.es/es/help/nnet/ref/trainlm.html>.
- [98] S. Sakai, K. Kobayashi, J. Nakamura, S. Toyabe y K. Akazawa, «Accuracy in the diagnostic prediction of acute appendicitis based on the bayesian network model» *Methods of Information in Medicine*, vol. 46, pp. 723-726, 2007.
- [99] S. Sayad, «An Introduction to Data Mining» [En línea]. Available: [http://www.saedsayad.com/support\\_vector\\_machine.htm](http://www.saedsayad.com/support_vector_machine.htm). [Último acceso: 2018].
- [100] J. Gibbons y S. Chakraborti, *Nonparametric Statistical Inference*, 5th Ed., Boca Raton: Chapman & Hall/CRC Press, Taylor & Francis Group, 2011.
- [101] M. Hollander y D. Wolfe, *Nonparametric Statistical Methods*, John Wiley & Sons, 1999.
- [102] T. Masters, «Practical Neural Networks recipes in C++» Academic Press, Inc., 1993, pp. 173-180.

- [103] «Matlab R2013a Online Documentation,» [En línea]. Available: <http://www.mathworks.es/es/help/nnet/ref/tansig.html>.
- [104] F. Medina y M. Galván, *Imputación de datos: teoría y práctica*, D. d. E. y. P. E. Naciones Unidas, Ed., Santiago de Chile: CEPAL, 2007.
- [105] S. Pita-Fernández y S. Pértegas-Díaz, «Pruebas diagnósticas: Sensibilidad y especificidad» *Cad Aten Primaria*, vol. 10, pp. 120-124, 2003.
- [106] S. Tortajada, J. Garcia-Gomez, J. Vicente, J. Sanjuan, R. Frutos, R. Martín-Santos, L. García-Esteve, I. Gornemann, A. Gutiérrez-Zotes, F. Canellas, A. Carracedo, M. Gratacos, R. Guillamat, E. Baca-García y M. Robles, «Prediction of Postpartum Depression Using Multilayer Perceptrons and Pruning» *Methods of Information in Medicine*, vol. 48, nº 3, pp. 291-298, 2009.
- [107] V. Fuster, L. Ryden, R. Asinger, D. Cannom, H. Crijns y R. Frye, «ACC/AHA/ESC guidelines for the management of patients with atrial fibrillation: executive summary» *Journal of the American College of Cardiology*, vol. 38, nº 4, pp. 1231-66, 2001.
- [108] A. Dobson, *An introduction to Generalized Linear Models*, New York: Chapman & Hall, 1990.
- [109] A. Burashnikov, J. DiDiego, A. Zygmunt, L. Belardinelli y C. Antzelevitch, «Atrium-selective sodium channel block as a strategy for suppression of atrial fibrillation: Differences in sodium channel inactivation between atria and ventricles and the role of ranolazine» *Circulation*, vol. 116, nº 13, pp. 1449-57, 2007.
- [110] D. Fedida, P. Orth, J. Chen, S. Lin, B. Plouvier, G. Jung, A. Ezrin y G. Beatch, «The mechanism of atrial antiarrhythmic action of RSD1235» *Journal of cardiovascular electrophysiology*, vol. 16, nº 11, pp. 1127-38, 2005.
- [111] I. Stiell, J. Roos, K. Kavanagh y G. Dickinson, «A multicenter, open-label study of vernakalant for the conversion of atrial fibrillation to sinus rhythm» *American Heart Journal*, vol. 159, nº 6, pp. 1095-101, 2010.
- [112] A. Bollmann, K. Sonne, H. Esperer, I. Toepffer y H. Klein, «Patients with persistent atrial fibrillation taking oral verapamil exhibit a lower atrial frequency on the ECG» *Annals of Noninvasive Electrocardiology*, vol. 7, nº 2, pp. 92-97, 2002.

## Apéndice A – Publicación

---

Las investigaciones descritas en este trabajo permitieron la publicación de [95]:

*S. Jiménez-Serrano, J. Yagüe, E. Simarro, C. Calvo, F. Castells, J. Millet «Atrial Fibrillation Detection Using Feedforward Neural Networks and Automatically Extracted Signal Features» *Computing in Cardiology (CinC) 2017*, 44:131-134. Rennes (France), 2017.*

En esta publicación se detalla una primera experimentación en la que se obtuvieron resultados similares a los presentados en este Trabajo Final de Máster.

A continuación, se añade dicho artículo.

# Atrial Fibrillation Detection Using Feedforward Neural Networks and Automatically Extracted Signal Features

Santiago Jiménez-Serrano<sup>1</sup>, Jaime Yagüe-Mayans<sup>1</sup>, Elena Simarro-Mondéjar<sup>1</sup>, Conrado J. Calvo<sup>1</sup>, Francisco Castells<sup>1</sup>, José Millet<sup>1</sup>

<sup>1</sup>Instituto ITACA, Universitat Politècnica de València, València, Spain

## Abstract

*Atrial Fibrillation (AF) is becoming an increasingly significant clinical matter as its prevalence keeps growing. Therefore, developing algorithms to accurately detect AF episodes from single-lead ECG recordings would benefit early and automatic diagnosis from monitoring devices. The aim of this study is to provide a Feedforward Neural Network (FFNN) classification model and assess its performance in order to discriminate short single-lead ECG registers among 4 groups: Normal (N), AF (A), Other rhythms (O) and noisy (~).*

*We extracted automatically 72 features derived from ventricular activity from each of the 8528 ECG records provided by the 2017 PhysioNet/Computing in Cardiology Challenge. Next, we performed a Feature Selection (FS) and a training/validation process through a grid search over a set of FFNN training parameters. We used a F1 scoring in order to assess the classification performance.*

*Results shown that filtering 50 features during the FS stage improved the initial classification performance from  $F1=0.70$  to  $F1=0.73$ . The following tuning of FFNN training parameters showed the best results during the 10-Fold Cross-Validation with  $F1=0.76$  ( $F1n=0.87$ ,  $F1a=0.78$ ,  $F1o=0.65$ ,  $F1p=0.45$ ) using 200 epochs,  $\alpha=0.7$ ,  $\beta=0.0$ , and one hidden layer made of 128 units. The final score on the test data was  $F1=0.77$ , demonstrating the robustness of the presented method.*

*Our strategy revealed promising classification scores using a robust validation approach. The resulting classification model is computationally low consuming during classification, hence is a good candidate to be implemented in wearable patient management systems.*

## 1. Introduction

The clinical importance of cardiac arrhythmias is increasing along with their incidence and prevalence mostly associated with population aging [1]. Among these diseases, atrial fibrillation (AF) stands out [2] since its growing trend is more significant and since in many

occasions it can be difficult to diagnose, due to paroxysmic behaviour and the absence of symptoms in some cases.

On the other hand, nowadays we are in a scenario where wearable devices are gaining great interest as monitoring devices in both research and clinical ambits [3]. However, automatic methods to provide a reliable diagnostic of AF using the ECG signals provided by wearable devices is still challenging, particularly if other normal or abnormal rhythms are also considered.

The aim of this study is to provide a classification model and assess its performance in order to discriminate short single-lead ECG registers among 4 groups: Normal (N), AF (A), Other rhythms (O) and noisy (~); in the context of the 2017 PhysioNet/Computing in Cardiology Challenge [4].

The strategy proposed in this work is based on automatic signal Feature Extraction (FE) derived from ventricular activity, plus a posterior Feature Selection (FS). Finally, those selected features were used in order to train and validate the classification performance of a set of different Feedforward Neural Networks (FFNN) through a tuning of training parameters.

Since the signals used share many characteristics with those offered by wearables devices, the resulting classification model could be a good candidate to be implemented in wearable patient management systems being as this approach is computationally low consuming during classification.

## 2. Materials

As database for this study we used the 8528 single-lead ECG registers provided by the competition. Deeper explanation of the database can be found in [4].

## 3. Methods

This section describes the signal feature extraction and selection processes, plus the model validation methodology used during this work. All these stages were carried on through MATLAB (2017a, The MathWorks).

### 3.1. Signal preprocessing

Provided signals were already band-pass filtered between 0.5Hz and 40Hz, so we performed an artefacts filtering as unique signal preprocessing stage. To do so, we used a 0.5 second sliding window in order to calculate aberrant maximum and minimum values, and sections surrounded by outliers were not considered for analysis purposes.

### 3.2. Scoring

2017 PhysioNet/Computing in Cardiology Challenge scoring rules are described in [4], where  $F1a$ ,  $F1n$ ,  $F1o$  and  $F1p$  are the individual score for A, N, O and  $\sim$  classes respectively.  $F1$  represents the mean among  $F1a$ ,  $F1n$  and  $F1o$  values, and consequently the classification performance for a given model in this work. We report  $F1p$  although is not necessary in order to calculate  $F1$ .

### 3.3. Feature Extraction

We automatically extracted 72 signal features derived from ventricular activity from each ECG record. Atrial activity related features were avoided since atrial signal can be easily obscured by artifacts and noise, or even not be present in the signal. To carry out this task, initially, we extracted the  $RR$  sequence using a  $QRS$  detector based on the first derivative of the ECG. Then we filtered the outliers from the  $RR$  sequence, and obtained the first and second derivatives of that sequence ( $RRd1$ ,  $RRd2$ ).

Using the above information, the extracted signal features can be grouped as:

1. Basic statistics (mean, standard deviation, kurtosis, skewness) over the  $RR$ ,  $RRd1$  and  $RRd2$  sequences. 9 features.
2. Features based on  $RRd1$ :  $RMSSD$ ,  $pNN25$ ,  $pNN50$ ,  $pNN75$ , where  $pNNxx$  denotes the percentage of intervals between normal beats exceeding  $xx$  ms. 4 features.
3. Lorenz plot-based features using  $RRd2$ : Angular variability, dispersion of the distance between points to origin, and differences between 2 and 3 consecutive beats. 8 features.
4. Same statistics as in points 1 and 2, but using an 8 seconds sliding window and a step size of 2 seconds. Once the matrix of values is obtained using each signal interval, we extracted the minimum, maximum, mean and standard deviation for each feature, appending all this values in a 44 features vector.
5. Other features: Shannon entropy of the  $RR$  sequence, Lempel-Ziv complexity of the  $RR$  time series after binarization using the median as threshold, and ratio between the number of different  $QRS$  patterns found and the total number of waves detected. 3 features.
6. Dynamic Time Warping ( $DTW$ ) distance value among the mean  $pNNx$  vector [5] for each class ( $pNNx(A)$ ,

$pNNx(N)$ ,  $pNNx(O)$ ,  $pNNx(\sim)$ ) and the corresponding  $pNNx(i)$  vector for each  $i$  sample. 4 features.

The last features group was the unique that needed a previous processing of the dataset in order to obtain the mean  $pNNx$  vectors for each class. Figures 1 and 2 show the result from varying the threshold, denoted by  $x$ , to obtain the ratio of consecutive  $RR$  that differs more than  $x$  ms. Figure 2 also shows standard deviation of measures for N and A groups.

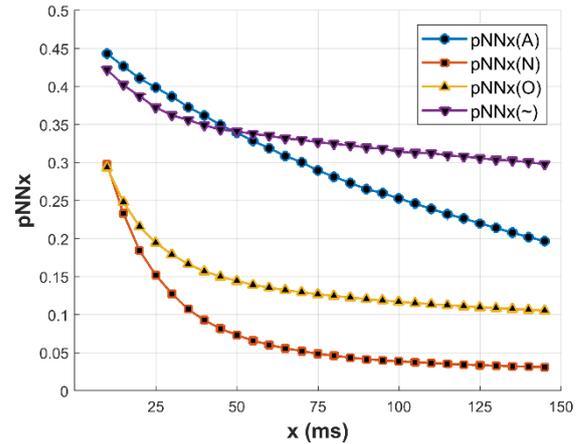


Figure 1: Mean  $pNNx$  vectors [5] obtained at the provided dataset for each class:  $pNNx(A)$ ,  $pNNx(N)$ ,  $pNNx(O)$  and  $pNNx(\sim)$ . These vectors are necessary in order to get the  $DTW$  distance of any other  $i$  sample with its own  $pNNx(i)$ .

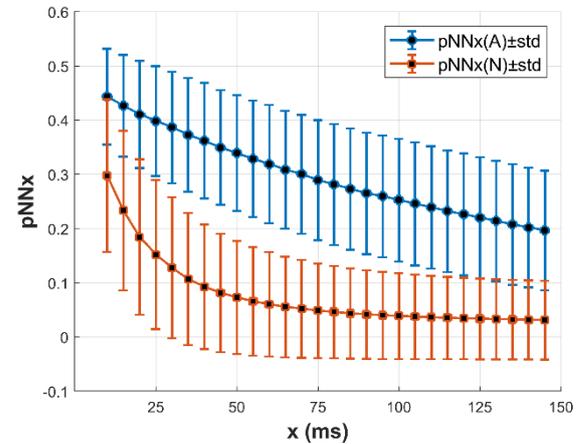


Figure 2: Mean  $pNNx$  vectors [5] and their standard deviation for class A and N ( $pNNx(A)$ ,  $pNNx(N)$ ).

Using the vectors presented above, we propose 4 new features (group 6) using the  $DTW$  metric in order to calculate a distance among the  $pNNx$  vector generated for each  $i$  sample ( $pNNx(i)$ ), and each mean  $pNNx$  vectors corresponding to each class:  $pNNx(A)$ ,  $pNNx(N)$ ,  $pNNx(O)$  and  $pNNx(\sim)$ . Hence, these new distance features are denoted as

$$\begin{aligned}
 DtwNNa(i) &= DTW(pNNx(A), pNNx(i)) \\
 DtwNNn(i) &= DTW(pNNx(N), pNNx(i)) \\
 DtwNNo(i) &= DTW(pNNx(O), pNNx(i)) \\
 DtwNNp(i) &= DTW(pNNx(\sim), pNNx(i))
 \end{aligned}$$

Figure 3 shows the boxplot using  $DtwNNn$  values in our dataset. Using all the above detailed features, we created a dataset consisting of 8528 samples and 72 features for each sample.

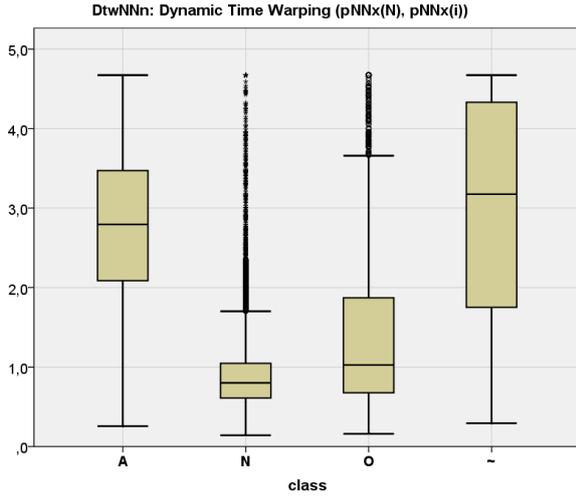


Figure 3: Boxplot of the  $DtwNNn$  proposed feature values obtained using the  $DTW$  distance among each  $pNNx(i)$  value of the database and the mean  $pNNx(N)$  vector corresponding to the Normal rhythm class.

### 3.4. Feature dataset preprocessing

First, for each feature, outliers exceeding 3 times the standard deviation above or below the median were replaced by these same limits.

Next, if some sample contained a  $NaN$  value due to a feature extraction error, we removed that sample from the dataset. According to this rule, 25 samples were removed from the dataset, obtaining a final dataset with 8503 samples. Since 64% of removed samples corresponded to Noisy class, during the test stage we classify as ( $\sim$ ) those containing some  $NaN$  value. Finally, we performed a z-score using the training set to rescale the whole dataset.

### 3.5. Feature Selection

A FFNN with 1 hidden layer of 128 units was used in order to determine the  $F1$  score through a hold-out strategy (50%-50% train/test split), using 75 epochs to avoid overfitting. To determine the best learning rate ( $\alpha$ ) and momentum ( $\beta$ ) FFNN training parameters for this stage, an initial grid-search was performed. Next, we carried out a Feature Selection (FS) using a backward-elimination

approach. Initially, all the variables are included in the dataset. A relaxation condition was added to the FS process in order to avoid early stopping criteria: if during the current iteration ( $j$ ) of the algorithm,  $F1$  score does not increase respect the ( $j-1$ ) iteration, we select as feature to be removed the one that maximized  $F1$  during the iteration ( $j$ ) if and only if  $F1_{(j)} > (F1_{(j-1)} - 0.05)$ .

### 3.6. FFNN training and validation

Using as inputs the selected features, we performed a grid search over the training parameters of a FFNN ( $\alpha$ ,  $\beta$ , architecture) in two stages: 1) A wide grid-search using a hold-out validation approach (50%-50% train/test split) using 100 epochs; 2) Assessment of the 10 best training parameters combinations through a 10-Fold Cross-Validation (10F-CV) using 100 and 200 epochs respectively. Finally, parameters allowing to achieve higher  $F1$  score were used in order to train the final FFNN with the whole dataset, to be submitted for the Challenge competition.

## 4. Results

Results showed an initial  $F1$  score of 0.709 using the 72 original features, achieved with  $\alpha=0.6$ ,  $\beta=0.1$  (FFNN 1 hidden layer of 128 units). During the FS stage, we removed 50 features, achieving a new  $F1$  score of 0.734. We found that the  $DtwNNn$  and  $DtwNNp$  proposed features were within the selected set of features.

Table 1. Best  $F1$  scores achieved during hold-out validation stage using 100 epochs. Training parameters: number of units in the hidden layer (Architecture), learning rate ( $\alpha$ ), momentum ( $\beta$ ).

Architecture	( $\alpha$ )	( $\beta$ )	$F1$
[1024]	0.7	0.1	0.7360
[512]	0.6	0.1	0.7356
[512]	0.5	0.1	0.7332
[128]	1.0	0.1	0.7320
[512]	0.5	0.4	0.7318
[128]	0.7	0.0	0.7310
[256]	0.5	0.1	0.7301
[256]	0.5	0.4	0.7301
[512]	0.8	0.1	0.7299
[512]	0.7	0.4	0.7298

Table 1 shows the results during the hold-out validation stage and Table 2 shows the results during the 10F-CV stage. All the FFNN reported consisted in 1 hidden layer. Best  $F1$  score value was 0.7644, achieved with a FFNN made of 118 units in the hidden layer, using 200 epochs,  $\alpha=0.7$  and  $\beta=0.0$ . Finally, the FFNN trained with the whole training dataset and the best training parameters achieve a final score value of  $F1=0.77$  on the test data.

Table 2. Best  $F1$  ( $F1a$ ,  $F1n$ ,  $F1o$ ,  $F1p$ ) scores achieved during the 10-Fold Cross-Validation stage.

#Epochs	Architecture	$(\alpha)$	$(\beta)$	$F1$	$F1a$	$F1n$	$F1o$	$F1p$
200	[128]	0.7	0.0	0.7644	0.7768	0.8668	0.6496	0.4490
200	[128]	0.6	0.1	0.7629	0.7761	0.8668	0.6458	0.4624
100	[512]	0.5	0.4	0.7627	0.7723	0.8645	0.6512	0.5112
100	[128]	0.7	0.0	0.7604	0.7766	0.8646	0.6402	0.4587
100	[512]	0.6	0.1	0.7603	0.7694	0.8637	0.6477	0.4844
200	[512]	0.5	0.1	0.7601	0.7656	0.8652	0.6494	0.4810
100	[1024]	0.7	0.1	0.7600	0.7649	0.8638	0.6513	0.3974
100	[512]	0.5	0.1	0.7599	0.7754	0.8632	0.6410	0.5097
100	[512]	0.8	0.1	0.7585	0.7681	0.8646	0.6429	0.4962
200	[1024]	0.7	0.1	0.7580	0.7629	0.8653	0.6457	0.4795
100	[512]	0.7	0.4	0.7578	0.7681	0.8593	0.6460	0.4959
200	[512]	0.5	0.4	0.7575	0.7635	0.8618	0.6471	0.4680

## 5. Discussion

Results obtained in this work showed differences in simple ventricular activity derived features for different rhythms, even when the arrhythmia had a supraventricular origin (AF). This allows for avoiding problems associated with absent or masked atrial activity. Also, FFNN showed good performance taking profit of those differences in order to classify the records in the 4 considered classes. Furthermore, the results of this method shows a high robustness since best F1 score in training and test data were 0.76 and 0.77 respectively, showing no overfitting of the model.

However, there were difficulties when discriminating O and ~ groups, since these groups aggregate records with different rhythms and characteristics, which results in great variability in considered features. Given this, we believe that including noise related features and further sub-grouping are strategies that could enhance the obtained results.

## 6. Conclusion

We presented and evaluated a robust method for short single-lead ECG registers classification that combines ventricular feature extraction and selection, and a FFNN training and validation approach. The result is a low computationally consuming classification model suitable for wearable monitoring devices. Improving the identification of other rhythms other than normal or AF incorporating more specific features for these cases into the proposed method, is an interesting direction to explore in the future.

## Acknowledgements

This work has been funded in part by the *Spanish Ministry of Economy and Competitiveness*: Programa Promoción de

Empleo Joven e Implantación de la Garantía Juvenil 2014 (PEJ-2014-A-36316).

## References

- [1] Chow GV, Marine JE, Fleg JL (2012). Epidemiology of Arrhythmias and Conduction Disorders in Older Adults. *Clinics in Geriatric Medicine*, 28(4), 539–553.
- [2] Chugh SS, Havmoeller R, Narayanan K, Singh D, Rienstra, M, Benjamin EJ, Murray CJL (2014). Worldwide Epidemiology of Atrial Fibrillation: A Global Burden of Disease 2010 Study. *Circulation*, 129(8), 837–847.
- [3] Kumari P, Mathew L, Syal P (2017). Increasing trend of wearables and multimodal interface for human activity monitoring: A review. *Biosensors and Bioelectronics*, 90: 298-307.
- [4] Gari Clifford, Chengyu Liu, Benjamin Moody, Ikaro Silva, Qiao Li, Alistair Johnson, Roger Mark, "AF Classification from a Short Single Lead ECG Recording: the PhysioNet Computing in Cardiology Challenge 2017", *Computing in Cardiology (Rennes: IEEE)*, Vol 44, 2017.
- [5] Mietus JE, Peng C-K, Henry I, Goldsmith RL, Goldberger AL. (2002). The pNNx files: re-examining a widely used heart rate variability measure. *Heart*, 88: 378-380.

Address for correspondence.

Santiago Jiménez Serrano  
 Universitat Politècnica de València, Instituto ITACA, Edif. 8G  
 Camino de Vera s/n, 46022-València (Spain)  
[sanjiser@upv.es](mailto:sanjiser@upv.es)



## Apéndice B – Ranking *Physionet Computing in Cardiology Challenge 2017*

Tabla 15. Ranking de puntuaciones durante la última fase de la competición *Physionet Computing in Cardiology Challenge 2017*. Los valores de  $F_1$ ,  $F_{1n}$ ,  $F_{1a}$ ,  $F_{1o}$  y  $F_{1p}$  se obtuvieron utilizando la versión final V3 del etiquetado de muestras, y los 3.658 registros ECG que los organizadores todavía mantienen como conjunto de datos de test privado. El valor  $F_1(Tr)$  indica el rendimiento de clasificación en el conjunto de 8.528 muestras de entrenamiento.

Ranking	$F_1$	$F_{1n}$	$F_{1a}$	$F_{1o}$	$F_{1p}$	$F_1(Tr)$	1.º Autor	Título de la publicación
1	<b>0.83</b>	0.9087	0.8351	0.734	0.5042	0.95	Morteza Zabihi	<a href="#">Detection of Atrial Fibrillation in ECG Hand-held Devices Using a Random Forest Classifier</a>
1	<b>0.83</b>	0.9117	0.8128	0.7505	0.5671	0.97	Shenda Hong	<a href="#">ENCASE: an ENsemble CIASsifiEr for ECG Classification Using Expert Features and Deep Neural Networks</a>
1	<b>0.83</b>	0.9158	0.8225	0.7498	0.5239	0.97	Shreyasi Datta	<a href="#">Identifying Normal, AF and other Abnormal ECG Rhythms using a Cascaded Binary Classifier</a>
1	<b>0.83</b>	0.903	0.8547	0.7366	0.5622	0.89	Tomás Teijeiro	<a href="#">Arrhythmia Classification from the Abductive Interpretation of Short Single-Lead ECG Records</a>
5	<b>0.82</b>	0.9112	0.8191	0.7328	0.5661	0.87	Guangyu Bin	<a href="#">Detection of Atrial Fibrillation Using Decision Tree Ensemble</a>
5	<b>0.82</b>	0.909	0.8221	0.7319	0.5676	0.89	Martin Zihlmann	<a href="#">Convolutional Recurrent Neural Networks for Electrocardiogram Classification</a>
5	<b>0.82</b>	0.9107	0.825	0.7299	0.5646	0.96	Mohammed Baydoun	
5	<b>0.82</b>	0.9031	0.8203	0.731	0.5251	0.88	Zhaohan Xiong	<a href="#">Robust ECG Signal Classification for the Detection of Atrial Fibrillation Using Novel Neural Networks</a>
9	<b>0.81</b>	0.9089	0.8007	0.7224	0.507	0.96	Shivnarayan Patidar	<a href="#">Automated Detection of Atrial Fibrillation using Fourier-Bessel expansion and Teager Energy Operator from Electrocardiogram Signals</a>
9	<b>0.81</b>	0.8991	0.8136	0.7111	0.5341	0.90	Chen Yao	
9	<b>0.81</b>	0.9129	0.7768	0.7272	0.5579	0.89	Dawid Smoleń	<a href="#">Atrial Fibrillation Detection Using Boosting and Stacking Ensemble</a>

Ranking	$F_1$	$F_{1n}$	$F_{1a}$	$F_{1o}$	$F_{1p}$	$F_1(Tr)$	1.º Autor	Título de la publicación
9	<b>0.81</b>	0.9054	0.8021	0.736	0.543	0.93	Filip Plesinger	<a href="#">Automatic Detection of Atrial Fibrillation and Other Arrhythmias in Holter ECG Recordings using PQRS Morphology and Rhythm Features</a>
9	<b>0.81</b>	0.9086	0.8101	0.7037	0.4645	0.94	Marcus Vollmer	<a href="#">Can Supervised Learning Be Used to Classify Cardiac Rhythms?</a>
9	<b>0.81</b>	0.8997	0.8272	0.7178	0.6396	0.93	Martin Kropf	<a href="#">ECG Classification Based on Time and Frequency Domain Features Using Random Forrests</a>
9	<b>0.81</b>	0.9118	0.784	0.7393	0.5547	0.88	Lucia Billeci	<a href="#">Detection of AF and Other Rhythms Using RR Variability and ECG Spectrum Measures</a>
9	<b>0.81</b>	0.9037	0.8067	0.7157	0.5459	0.84	Radovan Smíšek	<a href="#">SVM Based ECG Classification Using Rhythm and Morphology Features, Cluster Analysis and Multilevel Noise Estimation</a>
9	<b>0.81</b>	0.9024	0.8156	0.7194	0.5707	0.97	Andrew Goodwin	<a href="#">Classification of Atrial Fibrillation Using Multidisciplinary Features and Gradient Boosting</a>
18	<b>0.80</b>	0.9085	0.7881	0.7177	0.5765	0.78	Chen Jiayu	
18	<b>0.80</b>	0.9031	0.7895	0.7076	0.5943	0.91	Dionisije Sopic	<a href="#">A Hierarchical Cardiac Rhythm Classification Methodology Based on Electrocardiogram Fiducial Points</a>
18	<b>0.80</b>	0.9	0.7846	0.7013	0.5469	0.87	Gliner Vadim	<a href="#">Identification of Features for Machine Learning Analysis for Automatic Arrhythmogenic Event Classification</a>
18	<b>0.80</b>	0.8989	0.8085	0.6974	0.5398	0.8	Ivaylo Christov	<a href="#">Multi-parametric Analysis for Atrial Fibrillation Classification in the ECG</a>
18	<b>0.80</b>	0.9034	0.8027	0.704	0.5274	0.84	Joachim A. Behar	<a href="#">Rhythm and Quality Classification from Short ECGs Recorded using a Mobile Device</a>
18	<b>0.80</b>	0.8932	0.8034	0.6969	0.5442	0.88	Saman Parvaneh	<a href="#">Densely Connected Convolutional Networks and Signal Quality Analysis to Detect Atrial Fibrillation Using Short Single-Lead ECG Recordings</a>
18	<b>0.80</b>	0.8993	0.7935	0.7035	0.4282	0.86	Philip A. Warrick	<a href="#">Cardiac Arrhythmia Detection from ECG Combining Convolutional and Long Short-Term Memory Networks</a>
25	<b>0.79</b>	0.8786	0.7953	0.6887	0.6404	0.87	Csaba Botos	
25	<b>0.79</b>	0.9062	0.7385	0.7165	0.4751	0.92	Fernando Andreotti	<a href="#">Comparing Feature Based Classifiers and Convolutional Neural Networks to Detect Arrhythmia from Short Segments of ECG</a>

Ranking	$F_1$	$F_{1n}$	$F_{1a}$	$F_{1o}$	$F_{1p}$	$F_1(Tr)$	1.º Autor	Título de la publicación
25	<b>0.79</b>	0.899	0.7708	0.6944	0.4121	0.85	Patrick Schwab	<a href="#">Beat by Beat: Classifying Cardiac Arrhythmias with Recurrent Neural Networks</a>
25	<b>0.79</b>	0.9125	0.757	0.7032	0.559	0.87	Sasan Yazdani	<a href="#">Heart Rhythm Classification using Short-term ECG Atrial and Ventricular Activity Analysis</a>
29	<b>0.78</b>	0.9013	0.744	0.6812	0.5983	0.87	C Maximilian Oremek	
29	<b>0.78</b>	0.8893	0.7785	0.681	0.628	0.96	Christoph Hoog Antink	<a href="#">Fusing QRS Detection, Waveform Features, and Robust Interval Estimation with a Random Forest to Classify Atrial Fibrillation</a>
29	<b>0.78</b>	0.9043	0.7452	0.6799	0.324	0.83	Nadi Sadr	
29	<b>0.78</b>	0.8766	0.7781	0.6702	0.5819	0.97	Yang Liu	<a href="#">Diagnosis of AF Based on Time and Frequency Features by using a Hierarchical Classifier</a>
29	<b>0.78</b>	0.9031	0.7508	0.6972	0.000	0.83	Vykintas Maknickas	<a href="#">Atrial Fibrillation Classification Using QRS Complex Features and LSTM</a>
34	<b>0.77</b>	0.8912	0.7673	0.6651	0.4368	0.89	Bradley M Whitaker	<a href="#">AF Classification from ECG Recording Using Feature Ensemble and Sparse Coding</a>
<b>34</b>	<b>0.77</b>	<b>0.8884</b>	<b>0.7647</b>	<b>0.6686</b>	<b>0.422</b>	<b>0.78</b>	<b>Jiménez-Serrano, Santiago</b>	<a href="#">Atrial Fibrillation Detection Using Feedforward Neural Networks and Automatically Extracted Signal Features</a>
34	<b>0.77</b>	0.872	0.7831	0.6571	0.5802	0.87	Miguel Lozano	
34	<b>0.77</b>	0.8923	0.7553	0.6715	0.4843	0.86	Mohamed Limam	<a href="#">AF Detection and ECG Classification based on Convolutional Recurrent Neural Network</a>
38	<b>0.76</b>	0.9005	0.7267	0.6445	0.5839	0.84	Teo Soo-Kng	
39	<b>0.75</b>	0.8875	0.7204	0.6551	0.5401	0.77	C Babak Afshin-Pour	
39	<b>0.75</b>	0.8943	0.7312	0.6365	0.5888	0.9	Pietro Bonizzi	<a href="#">Detection of Atrial Fibrillation Episodes from Short Single Lead Recordings by Means of Ensemble Learning</a>
39	<b>0.75</b>	0.8973	0.7012	0.642	0.4377	0.79	Katarzyna Stepień	<a href="#">Classification of ECG Recordings with Neural Networks Based on Specific Morphological Features and Regularity of the Signal</a>

Ranking	$F_1$	$F_{1n}$	$F_{1a}$	$F_{1o}$	$F_{1p}$	$F_1(Tr)$	1.º Autor	Título de la publicación
39	<b>0.75</b>	0.8672	0.7352	0.6445	0.5857	0.79	Rymko	
39	<b>0.75</b>	0.8928	0.7250	0.646	0.0465	0.94	Zhenning Mei	
44	<b>0.74</b>	0.8912	0.7010	0.6137	0.2215	0.75	Griet Goovaerts	
45	<b>0.73</b>	0.8667	0.7766	0.5368	0.5468	0.73	C Manuel García	<a href="#">Atrial Fibrillation Screening through Combined Timing Features of Short Single-Lead Electrocardiograms</a>
45	<b>0.73</b>	0.8826	0.6825	0.6239	0.5217	0.73	Heikki Väänänen	<a href="#">Electrocardiogram Classification -- a Human Expert Way</a>
45	<b>0.73</b>	0.8691	0.7482	0.5804	0.0313	0.91	Javier de la Torre Costa	
45	<b>0.73</b>	0.8798	0.7096	0.5906	0.4401	0.76	Pedro Álvarez	
45	<b>0.73</b>	0.8800	0.7430	0.5565	0.3109	0.70	Victor Manuel José Ocoa	
50	<b>0.72</b>	0.8766	0.6689	0.6272	0.3003	0.83	Vignesh Kalidas	
51	<b>0.71</b>	0.8570	0.7225	0.564	0.0763	0.94	Jos van der Westhuizen	
51	<b>0.71</b>	0.8664	0.7179	0.5463	0.4658	0.78	Shadi Ghiasi	<a href="#">Atrial Fibrillation Detection Using Feature Based Algorithm and Deep Conventional Neural Network</a>
53	<b>0.69</b>	0.8736	0.6072	0.602	0.119	0.89	Ahmad B. A. Hassanat	
54	<b>0.68</b>	0.8379	0.669	0.5194	0.4138	0.72	C Philip Aston	
55	<b>0.64</b>	0.8272	0.5961	0.5078	0.3591	0.68	Irena Jekova	<a href="#">Arrhythmia Classification via Time and Frequency Domain Analyses of Ventricular and Atrial Contractions</a>
56	<b>0.63</b>	0.8524	0.6088	0.4184	0.27	0.68	Lluís Borràs Ferrís	
57	<b>0.61</b>	0.811	0.6046	0.4278	0.5575	0.63	Ilya Potapov	

Ranking	$F_1$	$F_{1n}$	$F_{1a}$	$F_{1o}$	$F_{1p}$	$F_1(Tr)$	1.º Autor	Título de la publicación
57	<b>0.61</b>	0.8039	0.5446	0.4709	0.3464	0.86	Ruhallah Amandi M	
59	<b>0.58</b>	0.6589	0.6704	0.4247	0.4563	0.63	DA Silva-Filarder Matthieu	<a href="#">Combining Template-based and Feature-based Classification to Detect Atrial Fibrillation from a Short Single Lead ECG Recording</a>
60	<b>0.56</b>	0.8158	0.5863	0.2700	0.3981	0.56	María Rebeca Lliguin León	
61	<b>0.55</b>	0.8079	0.4009	0.4364	0.3680	0.57	Erin Coppola	<a href="#">Atrial Fibrillation Classification from a Short Single Lead ECG Recording Using Hierarchical Classifier</a>
62	<b>0.53</b>	0.8763	0.2642	0.4643	0.000	0.53	Carlos Fambuena Santos	
62	<b>0.53</b>	0.8234	0.3416	0.425	0.3188	0.55	Octavian-Lucian Hasna	<a href="#">Robust Feature Extraction from Noisy ECG for Atrial Fibrillation Detection</a>
64	<b>0.51</b>	0.7728	0.4048	0.3556	0.0000	0.62	Ines Chavarria Marques	
65	<b>0.50</b>	0.8287	0.3529	0.3181	0.1168	0.78	Mihalis Nicolaou	
66	<b>0.48</b>	0.6647	0.4462	0.3245	0.4960	0.49	C Sergey	
67	<b>0.25</b>	0.4448	0.1163	0.1818	0.0000	0.27	Raviteja Mullapudi	

