







# DetECCIÓN AUTOMÁTICA DE PACIENTES CON ARRITMIAS AURICULARES EN REGISTROS ELECTROCARDIOGRÁFICOS

Samuel Guerrero García, Sergio Andrés Romero Torres

Universidad Antonio Nariño  
Facultad de Ingeniería Mecánica, Electrónica y Biomédica  
Cartagena, Colombia  
2021



# Deteccción automática de pacientes con arritmias auriculares en registros electrocardiográficos

Samuel Guerrero García, Sergio Andrés Romero Torres

Proyecto de grado presentado como requisito parcial para optar al título en ingeniería  
biomédica

Director:

Ph.D. María Isabel Arteaga Orozco

Línea de Investigación:

Procesamiento de señales biomédicas

Universidad Antonio Nariño

Facultad de Ingeniería Mecánica, Electrónica y Biomédica

Cartagena, Colombia

2021



*"La presente investigación se la dedico a Galo Guerrero Niño y Patricia Cecilia García Altamar, quienes me han apoyado durante todo este proceso de formación personal y profesional hasta el día de hoy."*

***Samuel Guerrero García***

*"Dedico este logro a Dios, que siempre ilumina mi camino y salvaguarda mis pasos, a mis padres Sergio Romero y Shirley Torres que son los artífices de este logro, por creer en mí, en mi desarrollo, recibiendo de ellos los mejores valores , apoyo incondicional , y a todas las personas que me acompañaron en este camino de aprendizaje hasta finalizar mi meta  
Gracias"*

***Sergio Andrés Romero Torres***

"Muchos de los fracasos de la vida son de personas que no se dieron cuenta cuán cerca estaban del éxito cuando se dieron por vencidos."

***Thomas A. Edison***

# AGRADECIMIENTOS

Primeramente a Dios, quien me ha dado todas las habilidades con las que cuento para superar los diferentes obstáculos que se presentan en la vida y sacar el mejor provecho de cada situación difícil. Agradezco a mis padres quienes me han permitido a través de sus esfuerzos, tener una excelente formación académica y sobre todo personal, inculcándome valores y principios desde que tengo memoria, siempre brindándome todo el amor, la motivación y compañía que me han hecho ser el hombre de bien que soy hoy.

A mis hermanos Ana Cristina Guerrero García y Christian Andres Guerrero García, quienes serán mis amigos incondicionales de por vida. Siempre me han acompañado y compartido sus conocimientos profesionales durante mi proceso de formación profesional.

*Samuel Guerrero García*

En primer lugar, quiero agradecer a Dios por escuchar nuestras oraciones y estar atento en todo momento, a Jader Giraldo, quien nos orientó con su experticia académica en cada etapa de nuestro trabajo y así, poder conseguir los resultados finales. A mis hermanos Sebastián Romero, Santiago Romero, por apoyarme aun cuando mis ánimos decaían y que siempre han estado ahí para darme palabras de apoyo. Muchas gracias a todos.

*Sergio Andrés Romero Torres*



# Resumen

El presente trabajo investigativo tiene como propósito la detección automática de pacientes con arritmias auriculares usando registros electrocardiográficos, mediante la extracción de características y métodos de inteligencia artificial. Teniendo en cuenta la importancia del análisis de señales biomédicas en el desarrollo de tecnologías para prevenir, diagnosticar y tratar enfermedades cardiovasculares, las cuales son la principal causa de muerte en el mundo y en el departamento de Bolívar, se desarrolla un software en el entorno MATLAB, utilizando bases de datos. Las bases de datos seleccionadas corresponden a 49 registros ECG de arritmias auriculares, pertenecientes a pacientes del Hospital Universitario de Valencia en España y provistos por un investigador local. Además, 50 registros son de pacientes con ritmo sinusal normal correspondientes al grupo de control tomado de la base de datos PTB de physionet. Para ambas bases de datos la frecuencia de muestreo es de  $1Khz$  y cada registro contiene 12 derivaciones tomando un valor de duración fijo de 12 segundos. Para cada registro es llevado a cabo una etapa de pre-procesamiento para eliminar fuentes de ruido indeseadas como el ruido asociado a la actividad muscular (EMG), ruido térmico y corrección de la banda base. En la siguiente fase los complejos QRS son detectados mediante la aplicación del algoritmo de Pan-Tompkins y por medio de ello se estima la variabilidad de la frecuencia cardiaca, la cual es cuantificada usando medidas de dispersión estadísticas tales como la desviación estándar, rango intercuartil, rango, media y la media de la desviación absoluta. Además, se extraen características espectrales basadas en el análisis de Fourier, frecuencia fundamental, densidad espectral de potencia, así como los coeficientes de detalle del tercer nivel de descomposición de la transformada wavelet, usando la wavelet madre db2. Para reducir la dimensionalidad, el algoritmo de análisis de componentes principales es usado y se eligen los 5 descriptores más significativos, que servirán de entrada como predictores a herramientas de inteligencia artificial. Bajo el entorno del toolbox de inteligencia artificial de

MATLAB se evalúa el rendimiento de diferentes modelos de clasificación, alcanzando rangos de exactitud entre 91 % – 95 % en condiciones específicas.

# Abstract

The purpose of this research work is the automatic detection of patients with atrial arrhythmias in The purpose of this research work is the automatic detection of patients with atrial arrhythmias using electrocardiographic records, by means of feature extraction and artificial intelligence methods. Taking into account the importance of the analysis of biomedical signals in the development of technologies to prevent, diagnose and treat cardiovascular diseases, which are the main cause of death in the world and in the department of Bolivar, a software is developed in the MATLAB environment, using databases. The selected databases correspond to 49 ECG records of atrial arrhythmias, belonging to patients from the University Hospital of Valencia in Spain and provided by a local researcher. In addition, 50 records are from patients with normal sinus rhythm corresponding to the control group taken from the physionet PTB database. For both databases the sampling rate is  $1Khz$  and each recording contains 12 leads taking a fixed duration value of 12 seconds. For each recording a pre-processing stage is carried out to eliminate unwanted noise sources such as noise associated with muscle activity (EMG), thermal noise and baseband correction. In the next phase the QRS complexes are detected by applying the Pan-Tompkins algorithm and by means of this the heart rate variability is estimated, which is quantified using statistical dispersion measures such as standard deviation, interquartile range, range, mean, median and mean absolute deviation. In addition, spectral features are extracted based on Fourier analysis, fundamental frequency, power spectral density, as well as the detail coefficients of the third level of wavelet transform decomposition, using the mother wavelet db2. To reduce the dimensionality, the principal component analysis algorithm is used and the 5 most significant descriptors are chosen, which will serve as input as predictors to artificial intelligence tools. Under the MATLAB artificial intelligence toolbox environment, the performance of different classification models is evaluated, reaching accuracy ranges between 91 %–95 % under specific

conditions.

# Índice general

<b>AGRADECIMIENTOS</b>	<b>II</b>
<b>Resumen</b>	<b>IV</b>
<b>1. Introducción</b>	<b>1</b>
1.1. Objetivo general . . . . .	3
1.2. Objetivos específicos . . . . .	3
1.3. Antecedentes y estado del arte . . . . .	3
<b>2. Marco Teórico</b>	<b>8</b>
2.1. Electrocardiograma . . . . .	8
2.2. Sistema eléctrico cardíaco . . . . .	9
2.3. Procesamiento de señal ECG . . . . .	9
2.3.1. Pre-procesamiento . . . . .	10
2.3.2. Algoritmo de Pan-tompkin . . . . .	10
2.3.3. Análisis espectral . . . . .	11
2.3.4. Transformada de Fourier . . . . .	12
2.3.5. Transformada Discreta de Fourier . . . . .	12
2.3.6. Transformada Rápida de Fourier . . . . .	13
2.3.7. Transformada de wavelet . . . . .	13
2.3.8. Transformada de wavelet continua . . . . .	13
2.3.9. Transformada discreta de wavelet . . . . .	14
2.4. Medidas de dispersión . . . . .	14
2.4.1. Rango . . . . .	14
2.4.2. Rango intercuartil . . . . .	15
2.4.3. Desviación estándar . . . . .	15
2.4.4. Varianza . . . . .	15
2.4.5. Media de la desviación absoluta . . . . .	16
2.5. Métodos de inteligencia artificial . . . . .	16
2.5.1. Machine learning . . . . .	17
2.5.2. Big data . . . . .	17
2.6. Análisis de Componentes Principales . . . . .	18
2.6.1. Eigenvalues . . . . .	18
2.6.2. Eigenvectors . . . . .	19
2.6.3. Cálculo de las componentes principales . . . . .	19
<b>3. Metodología</b>	<b>20</b>

3.1. Pre-procesamiento . . . . .	22
3.2. Detección de complejos QRS usando el algoritmo de pan tompkins . . . . .	23
3.2.1. Regla de Decisión . . . . .	23
3.2.2. Marca Fiducial . . . . .	23
3.2.3. Umbral . . . . .	23
3.2.4. Búsqueda de los complejos QRS faltantes . . . . .	24
3.2.5. Eliminación de detecciones múltiples dentro del periodo refractorio . . . . .	25
3.2.6. Discriminación de la onda T . . . . .	25
3.3. Estimación de la variabilidad de la frecuencia cardíaca usando medidas de dispersión . . . . .	26
3.4. Creación del dataset . . . . .	26
3.5. Aplicación de los métodos de inteligencia artificial . . . . .	27
<b>4. Resultados</b>	<b>28</b>
4.1. Pre-procesamiento . . . . .	28
4.2. Detección de complejos QRS . . . . .	29
4.3. Extracción de características . . . . .	30
4.4. Inteligencia artificial . . . . .	33
4.4.1. Caso 1 . . . . .	33
4.4.2. Caso 2 . . . . .	37
4.4.3. Caso 3 . . . . .	41
4.4.4. Caso 4 . . . . .	45
<b>5. Análisis de resultado</b>	<b>50</b>
<b>6. Conclusiones y recomendaciones</b>	<b>53</b>
<b>Referencias</b>	<b>55</b>
<b>Apéndices</b>	<b>60</b>

# Índice de cuadros

5.0.1.especificidad y sensibilidad de cada caso, a continuación se relacionan los acrónimos. SE=sensibilidad, SP=especificidad, MD=medidas de dispersión, DSP=densidad espectral de potencia, FD=frecuencia dominante, WC=coeficientes wavelet, PCA=análisis de componentes principales . . . . .	51
---	----

# Índice de figuras

2.1.1. Señal ECG . . . . .	9
3.0.1. Diagrama de bloques para el desarrollo del software. . . . .	22
4.1.1. ECG de fibrilación auricular contenido en el archivo x003.mat, sin corrección de línea base y con ruido producido por actividad muscular. . . . .	28
4.1.2. ECG de fibrilación auricular contenido en el archivo x003.mat, luego de aplicar filtro butterworth. . . . .	29
4.2.1. Detección de complejos QRS en ECG de ritmo sinusal normal a través del algoritmo de Pan-Tompkins. . . . .	29
4.2.2. Detección de complejos QRS en ECG de fibrilación auricular a través del algoritmo de Pan-Tompkins. . . . .	30
4.2.3. Detección de complejos QRS en ECG de flutter auricular a través del algoritmo de Pan-Tompkins. . . . .	30
4.3.1. Gráfico comparativo para la desviación estandar entre los ECG. . . . .	31
4.3.2. Gráfico comparativo para la media de la desviación absoluta entre los ECG. . . . .	31
4.3.3. Gráfico comparativo para el rango entre los ECG. . . . .	32
4.3.4. Gráfico comparativo para el rango intercuartil entre los ECG. . . . .	32
4.3.5. Gráfico comparativo para la mediana entre los ECG. . . . .	33
4.4.1. Resultado obtenido para el modelo de clasificación Cubic SVM. Presenta una precisión de 95 %. . . . .	34
4.4.2. Gráfica de dispersión para el caso 1. . . . .	34
4.4.3. Coordenadas paralelas para el caso 1. . . . .	35
4.4.4. Matriz de confusión para el caso 1. . . . .	36
4.4.5. Región de convergencia para el caso 1 . . . . .	37
4.4.6. Resultado obtenido para el modelo de clasificación Linear SVM. Presenta una precisión de 92 %. . . . .	37
4.4.7. Gráfico de dispersión para el caso 2. . . . .	38
4.4.8. Coordenadas paralelas para el caso 2. . . . .	39
4.4.9. Matriz de confusión para el caso 2. . . . .	40
4.4.10. Región de convergencia para el caso 2. . . . .	41
4.4.11. Resultado obtenido para el modelo de clasificación Linear SVM. Presenta una precisión de 94 %. . . . .	42
4.4.12. Gráfico de dispersión para el caso 3. . . . .	42
4.4.13. Coordenadas paralelas para el caso 3. . . . .	43
4.4.14. Matriz de confusión para el caso 3. . . . .	44
4.4.15. Región de convergencia para el caso 3. . . . .	45



4.4.1	Resultado obtenido para el modelo de clasificación Linear SVM. Se obtuvo una precisión de 94 % . . . . .	46
4.4.17	Gráfico de dispersión para el caso 4. . . . .	46
4.4.18	Coordenadas paralelas para el caso 4. . . . .	47
4.4.19	Matriz de confusión para el caso 4. . . . .	48
4.4.20	Región de convergencia para el caso 4. . . . .	49

# Capítulo 1

## Introducción

Las enfermedades cardiovasculares son una de las causas de muerte más comunes a nivel mundial, según la Organización Mundial de la Salud (OMS), en el año 2000 se registraron 2 millones de defunciones y 8,9 millones en el año 2019, todas correspondientes a cardiopatía isquémica [1]. Según el Departamento Administrativo Nacional de Estadística (DANE), en el 2019, hubo un registro en Colombia de 35.323 defunciones causadas por infarto agudo de miocardio, de las cuales 1.021 corresponden al departamento de Bolívar [2]. Dentro de las arritmias cardiacas la más común es la Fibrilación auricular (FA) que afecta hasta 2% de la población mundial y es causada debido a un impulso eléctrico irregular en el nodo sinusal. La FA tiene importantes factores de riesgo asociados como el accidente cerebrovascular y la insuficiencia renal crónica [3]. Para prevenir, diagnosticar y tratar enfermedades cardiovasculares, los avances tecnológicos han permitido desarrollar herramientas tales como, el Big Data (BD) y la Inteligencia Artificial (IA) [4]. Entre las ventajas brindadas por estos avances en la medicina, destaca la oportunidad que le permite a los especialistas cardiovasculares realizar diagnósticos de manera más eficiente y eficaz. Por ejemplo, la IA y el BD, permiten almacenar y procesar datos a nivel informático, en este caso ECG, así los cardiólogos pueden observar, dar diagnósticos e iniciar tratamientos oportunos sobre las patologías de sus pacientes, como arritmias cardiacas, cardiopatía isquémica, entre otras. Lo anterior, se da accediendo desde un computador o dispositivo móvil, que a través de una red local o inalámbrica, se conecte a la base de datos que contenga los registros ECG obtenidos de los pacientes [5].

La técnica de diagnóstico de enfermedades cardiovasculares más utilizada es el ECG, que es un registro eléctrico de la actividad del corazón. El ECG se compone de diferentes ondas que representan cada una de las fases de los latidos del corazón, entre ellas se encuentra, onda P que representa la contracción auricular, complejo QRS que es la componente de mayor energía y representa la contracción ventricular. Finalmente, la onda T y U que representan la fase inicial y final de relajación ventricular. Diferentes métodos se han desarrollado para la detección y clasificación de cardiopatías por medio del análisis de señales ECG y el uso de herramientas de inteligencia artificial [6]. Dentro del concepto de inteligencia artificial, se encuentran las herramientas de machine learning, las cual se basa en algoritmos bioinspirados, tales como, redes neuronales cerebrales, algoritmo genético, enjambre de abejas, entre otros. Estos sistemas son entrenados por épocas en el cual las entradas o predictores son presentadas al algoritmo varias veces hasta que este alcanza su convergencia aprendiendo un modelo que se ajusta al comportamiento del fenómeno que se describe. Para mejorar el desempeño de los métodos de machine learning es necesario una etapa que contemple la extracción de características de las señales de entrada. La extracción de características permite extraer información relevante del evento que se quiere detectar, es importante que las características seleccionadas sean dicientes y no redundantes. Para evitar la redundancia es posible emplear algoritmos que permitan la reducción de dimensionalidad de los datos de entrada por medio de los valores propios del mismo [7]. La metodología de este trabajo contempla 3 etapas: pre-procesamiento, extracción de características e inteligencia artificial para la toma de decisiones. En la primera etapa un filtro pasa alta con frecuencia de corte de 1Hz es implementado para eliminar ruido de baja frecuencia. Para la extracción de características se hace uso del algoritmo de Pan Tompkins [8], en orden de detectar los complejos QRS. A partir de la detección de los complejos QRS la variabilidad de la frecuencia cardiaca es estimada y cuantificada en base a diversas medidas de dispersión, tales como, desviación estándar, rango intercuartil, rango, media y media de la desviación absoluta. Para completar el set de características, características espectrales asociadas a la transformada de Fourier (TF), tales como, densidad espectral de potencia y la frecuencia fundamental fueron extraídas, así como los coeficientes del tercer nivel de descomposición wavelet con la wavelet madre db2. Los resultados arrojan niveles de exactitud en el rango de 95% - 100% distinguiendo pacientes con arritmias auriculares FA, alteo auricular (AFL) y ritmo sinusal normal.

## 1.1. Objetivo general

- Detectar de forma automática arritmias auriculares en registros electrocardiográficos de pacientes mediante la extracción de características y métodos de inteligencia artificial.

## 1.2. Objetivos específicos

- Definir las técnicas para el pre-procesamiento de la señal de ECG, es decir, filtrado y acondicionamiento.
- Implementar análisis espectral y medidas estadísticas para la extracción de características de los registros ECG.
- Contrastar el desempeño de las técnicas de extracción de características usando modelos de machine learning.

## 1.3. Antecedentes y estado del arte

El ECG, permite observar la electrofisiología del corazón, mostrando distintas características funcionales del mismo, estas características permiten el diagnóstico y detección de diferentes patologías asociadas al corazón, entre ellas, las arritmias cardiacas. De acuerdo con lo anterior las características de las señales electrocardiográficas son extraídas en orden de encontrar relaciones entre las mismas con eventos de interés que desean ser detectados [9].

En un proyecto realizado por Cristian Vidal Silva y Leopoldo Pavesi Farriol, denominado “desarrollo de un sistema de adquisición y tratamiento de señales electrocardiográficas”[10], se llevó a cabo la construcción de un electrocardiógrafo digital de bajo costo y la implementación de algoritmos para medir los parámetros relevantes al diagnóstico final. El sistema consta de las siguientes etapas: etapa de adquisición de biopotenciales cardiacos, etapa de amplificación, etapa de filtrado y la etapa de adecuación de datos. Logrando tener como resultado una buena calidad de producto, teniendo en cuenta el bajo costo de los componentes, demostrando que el sistema de adquisición de datos puede ser utilizado para la práctica. De igual forma, el proyecto denominado “extracción del ECG fetal (FECG) en electrocardiografía no invasiva por medio de señal de referencia”[11], desarrollado en la Universidad Julio Garavito Bogotá, Colombia por Ana Vivian Calderón Echeverría, donde se diseñó un algoritmo para la

extracción de la FEKG y de la frecuencia cardíaca fetal a partir de electrocardiografía no invasiva. La base de datos utilizada es Physionet donde extrajeron registros que contenían cuatro canales abdominales, a los cuales se le aplicaron técnicas de procesamiento. Este procedimiento se basa en el uso de filtro butterworth de orden cuatro, filtro lineal Wiener, análisis de componentes independiente y detección de complejo QRS por medio del algoritmo de Pan-Tompkins. Los resultados arrojan un 74 % de efectividad para una base de datos de 69 registros aplicables, considerando una aceptación del 5 % de error en la medida.

La tesis de pregrado “procesamiento del complejo QRS característico del electrocardiograma (ECG)” [12], desarrollada en la universidad de Málaga, por Nissrin Amrani El Yaakoubi, donde se realiza una evaluación del algoritmo de Pan - Tompkins que comprende dos finalidades, verificar el correcto funcionamiento de este explicando cada una de las fases que lo compone y analizar los resultados de la evaluación en busca de fragmentos de código mejorables en señales ECG. De esta forma, se pueda comprender las fases del algoritmo y los errores que este puede presentar y las mejoras propuestas en el código del algoritmo. Este estudio encontró que cuanto más parametrizados estén los procesos, mayor es la estabilidad funcional en diferentes tipos de señales ECG.

Muchas investigaciones se interesan por la clasificación de las anomalías cardíacas mediante el uso de patrones de la señal ECG. El registro ECG refleja la actividad eléctrica del corazón, cualquier cambio que se produzca en los patrones de la forma de onda del ECG en un tiempo prolongado puede indicar un síntoma asociado a patologías en el funcionamiento del corazón. En el análisis de la señal ECG con algoritmos precisos puede ayudar a diagnosticar los trastornos en una fase temprana. También, se puede inferir de la investigación titulada “Reducción del ruido en señales electrocardiográficas mediante la Transformada Wavelet”[13], que existen diferentes métodos para la reducción de ruido en señales ECG a través de la transformada de wavelet (WT), debido a que esta funciona por medio de banco de filtros. Se da uso de las diferentes wavelets madre y niveles de descomposición dependiendo del caso en estudio. Por último, es importante mencionar la utilidad de la WT para la reducción de la variación de la línea base, lo cual se implementó en el presente trabajo como uno de los pasos previos para posteriormente extraer componentes espectrales del complejo QRS y clasificarlos para la detección de arritmias. En el estudio de Montes V et al.[14] se presenta una investigación en la cual se utilizó la WT, análisis no lineal y mediciones de diagnóstico, para

extraer características de señales ECG con cardiopatía isquémica y normales. Posteriormente, se comparó el porcentaje de error entre las técnicas de extracción empleadas con el fin de evaluarlas a través de análisis discriminante lineal y máquinas de soporte vectorial (SVM). Finalmente, se obtiene mediante el análisis discriminante lineal que, las características más efectivas son extraídas a través de análisis no lineal de componentes principales sobre un latido cardíaco. El error de clasificación fue de un 0,22 %, el cual es menor en comparación con la WT, 6,78 %, y las mediciones de diagnóstico en un 24,22 %. También se observa que, con las máquinas de soporte vectorial, el resultado para WT tiene una precisión de hasta 0,1 %, para el de diagnóstico 5,11 % y 0,12 % para el análisis no lineal. Sin embargo, se concluyó que la WT es la más apropiada para el proceso de extracción de características de la funcionalidad cardíaca debido a que elimina el ruido con mayor efectividad.

La tesis de pregrado “Clasificación de arritmias cardíacas empleando la transformada wavelet y redes neuronales” [15], desarrollada en la Universidad tecnológica de Bolívar , por Harold Blanco Ortega, desarrolló un sistema de identificación de patologías del corazón, para lo que se ejecutó un algoritmo basado en la transformada wavelet continua. Lo anterior, es de utilidad al extraer las diferentes características de la señal ECG, obteniendo los componentes en frecuencia de la señal, evitando así eliminar algunos que pueden ser de vital importancia para el estudio de la misma. Para finalizar, se plantea el uso de un clasificador, el cual trata de redes neuronales, como estrategia para la etapa de clasificación de arritmias.

En el trabajo desarrollado en el 2007, en la Universidad Tecnológico de Pereira, por Wilmer h. Rojas, Jhon f. Herrera, entre otros, titulado “Análisis y supervisión de la señal cardíaca con transmisión inalámbrica de datos”[16], se describe el diseño y la implementación de un prototipo de transmisión vía RF, para el análisis y supervisión de la señal cardíaca. Dicho prototipo se basa de un electrocardiógrafo portátil con transmisión inalámbrica y una herramienta computacional para la visualización y análisis de la señal obtenida de las derivaciones DI, DII y DIII, utilizando el método wavelets para eliminar componentes ajenas a la señal electrocardiográfica y métodos basados en derivadas para extraer las características propias de cada onda cardíaca que son introducidas a una SVM, para su posterior reconocimiento.

En “Técnicas de extracción de características de señales biomédicas” [17], proyecto de investigación desarrollado en 2007, por Ángela Cabarcas Barboza y Tania Guerrero Castilla,

donde se realiza una revisión de las técnicas de extracción que pueda ser aplicada a las distintas señales biomédicas, dentro de las cuales se destacan los métodos clásicos basados en la TF y los métodos modernos como lo son la WT y las Redes Neuronales.

La investigación realizada por Anas, E et al.[18], propone un algoritmo de detección secuencial para clasificar patologías cardíacas graves, de otras arritmias. Como resultado se detecta correctamente las señales de taquicardia ventricular (VT) y fibrilación ventricular (VF) de otras arritmias, la precisión que se obtuvo fue de 99,32 %. También, se observa que algunas señales son muy difíciles de clasificar incluso por cardiólogos. Por lo tanto, la precisión de la técnica aplicada en la investigación presenta fallos ligeros debido a estos episodios confusos. En el trabajo “Análisis de señal del impulso cardíaco para el mejoramiento del diagnóstico de patologías del corazón” [19], desarrollado en 2011 , establece el análisis y comparación sobre varias técnicas para el análisis de señales cardíacas, donde se utiliza WT, dicho programa permite identificar la magnitud del QRS, lo cual ayuda en el diagnóstico de hipertrofia ventricular derecha e izquierda, y la duración del QT, cuya importancia radica en que su prolongación, puede predisponer a los pacientes a presentar taquicardia ventricular.

Por otro lado, Aya F. Khalaf [20], emplean la señal ECG para clasificar cinco tipos de latidos, entre los que se encuentran el normal, la contracción ventricular prematura, la contracción auricular prematura, el bloqueo de rama izquierda (LBBB) y el bloqueo de rama derecha (RBBB). Para ello, utilizaron un algoritmo basado en la transformación estadística no lineal de los datos para investigar las características ocultas de la señal. En su investigación, se empleó SVM para clasificar los latidos del corazón. Se obtuvo una tasa de precisión del 98,60 %.

Se realizó una investigación por parte del ingeniero Fernando Riveros Sanabria[21], en la cual se construyó una base de datos con registros ECG de diferentes pacientes con cardiopatías y enfermedad de Chagas, pertenecientes al departamento del Meta en Colombia. Además, se elaboró un algoritmo que permitió detectar las características ECG de las muestras tomadas, a través de la función Symlets 7 de la WT con 11 niveles de descomposición, la igualación a cero los coeficientes de los niveles pertinentes y finalmente se reconstruyó la señal, permitiendo de esta forma el filtrado y la detección de características. Así pues, se obtuvo un programa que permite detectar características ECG de arritmias cardíacas, registros ECG normales y pacientes con enfermedad de Chagas con una precisión mayor a 98 %.

El trabajo titulado “An Approach for ECG Feature Extraction using Daubechies 4 (DB4) Wavelet” [22], presenta el uso de la transformada de wavelet de forma similar al trabajo anterior, ya que se implementa la familia Daubechies 4 para desarrollar un algoritmo que permite extraer características de señales ECG de la base de datos MIT-BIH y así contribuir al desarrollo de mejores herramientas de diagnóstico de cardiopatías. Principalmente se detectan los picos R luego de filtrar la señal. También, menciona que a través del uso de ventanas proporcionales a los valores estándar de los intervalos normales de los picos R, es posible detectar las ondas T y P, las cuales también son detectadas en el desarrollo del algoritmo. Como resultado final, se observó que el programa propuesto tiene un desempeño con un error de desviación menor del 10 %.

Asimismo, los resultados obtenidos en el trabajo titulado “Transformada wavelet y máquinas de soporte vectorial para la identificación de arritmias” [23], muestran que el porcentaje de error de validación es reducido en el proceso de clasificación de ritmo sinusal normal y 4 tipos de arritmias elegidas. al emplearse la SVM con función núcleo de base radial en conjunto con la técnica de Análisis de Componentes Principales (PCA) se observa un buen desempeño de la WT en la extracción de características. Diferentes wavelets madres han sido usadas cada una con 6 niveles de descomposición.

Por otro lado, en la investigación titulada “Cálculo del umbral para detección de la onda R del complejo cardiaco” [24], realizada el 2014 por el ingeniero J. E. González-Barajas, se aplicó un banco de filtros para realizar un análisis de histograma. Para esta investigación un rango de frecuencias en específico es seleccionado en uno de los filtros para formar el histograma. Por medio de una matriz compuesta por distintos valores de amplitud y la frecuencia de ocurrencia de los mismos, se derivó dos veces el histograma y dio como resultado un valor de umbral para la detección de la onda R.

Finalmente, en el documento “Automatic detection of cardiac arrhythmia through ECG signal analysis: a review” [25], se lleva a cabo un amplio estudio sobre varios detalles ECG y la detención de anomalías cardiacas. Además, se propone un algoritmo que encuentra el complejo QRS basado en el doble criterio de la amplitud y la duración del complejo QRS. El algoritmo consiste en operaciones computacionales a través de la WT, que permite detectar las complejos QRS así como las ondas P y T.



## Capítulo 2

# Marco Teórico

### 2.1. Electrocardiograma

El electrocardiograma (ECG) representa la actividad eléctrica de las células del corazón. El impulso eléctrico generado en el nodo sinusal, produce la contracción rítmica del corazón. De igual manera, esta actividad electromecánica se ocasiona según un orden preciso y periódico. Medidas simples y compuestas del ECG se utilizan generalmente para detectar y clasificar diferentes ritmos cardiacos [26]. El ECG está compuesto por un conjunto de ondas que Einthoven denominó P, Q, R, S, T y U de acuerdo con el orden de aparición en el tiempo.

Se debe tener en cuenta que la onda P se inscribe como resultado de la activación auricular y se encuentra precedida por los complejos Q, R y S, los cuales representan la propagación de la onda de excitación a la musculatura de ambos ventrículos y al tabique interventricular. Terminando el proceso de despolarización de toda la masa muscular auricular y ventricular, acontece una pequeña pausa (que luego conoceremos como segmento S-T) y más tarde se inscribe la onda T que corresponde al inicio de la fase de relajación ventricular[27]. Ahora bien, las ondas son las distintas curvaturas que toma un trazo de ECG bien sea hacia arriba o hacia abajo en base a la proyección de cada una de las derivaciones. Estos, son productos de los potenciales de acción registrados durante la estimulación cardíaca vistos desde diferentes ángulos. La figura 2.1.1 muestra un ECG típico.

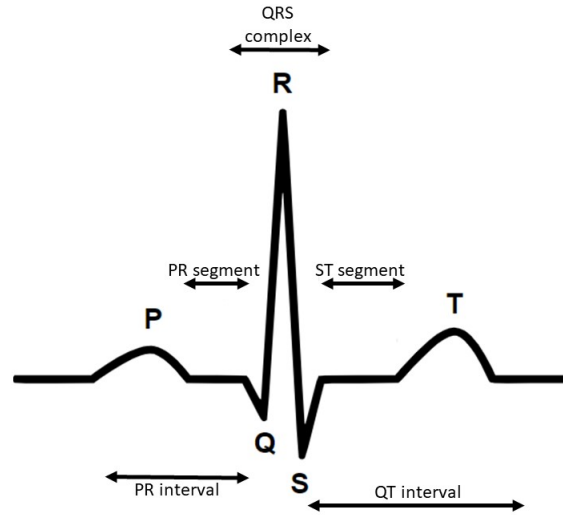


Figura 2.1.1: Señal ECG

## 2.2. Sistema eléctrico cardíaco

Es necesario saber que el corazón tiene un sistema de creación y envío de los impulsos eléctricos, los cuales son denominados señales eléctricas. En primer lugar, estos impulsos hacen que se contraigan las aurículas (las dos cámaras superiores del corazón), con lo cual la sangre pasa a los ventrículos (las dos cámaras inferiores del corazón). Luego, los impulsos eléctricos hacen que se contraigan los ventrículos, con lo cual la sangre es bombeada hacia los pulmones y el resto del cuerpo [28]. Ahora bien, las células del nodo sinoauricular de las cámaras superiores (las aurículas) en el músculo cardíaco (el miocardio) de forma espontánea y periódica cambian su polarización, que se extiende progresivamente a todo el músculo cardíaco. Esta despolarización eléctrica periódica y progresiva del miocardio se registra como pequeñas desviaciones potenciales entre dos lugares diferentes del cuerpo humano o con respecto a un electrodo de referencia. Una señal eléctrica casi periódica [29].

## 2.3. Procesamiento de señal ECG

Las técnicas utilizadas para el procesamiento de señales y las representaciones de tiempo-frecuencia, que sirven para poder obtener adecuadamente la máxima información de un ECG, tales como la transformada Fourier y transformada de wavelet, son usadas como principio para el diagnóstico automático de algún tipo de episodio de arritmias. Estas técnicas consisten en

extraer características en el dominio del tiempo o en el dominio de la frecuencia. Algunas de estas técnicas comparten similitudes tales como: combinar información temporal y espectral en la misma representación, lo cual es de vital importancia en las señales no estacionarias como las del ECG, especialmente cuando aparecen patologías irregulares como las arritmias cardiacas. Las técnicas de detección emplean típicamente parámetros obtenidos de la fase de pre-procesamiento de la señal para alimentar la entrada de un algoritmo de clasificación, esto se traduce en una pérdida de información de dicha señal. Esta información perdida puede ser clave para discriminar adecuadamente entre distintas patologías cardiacas. Por lo tanto, el desarrollo e implementación de un algoritmo que permita discriminar con mayor precisión entre las diversas condiciones cardiacas es fundamental [30].

### 2.3.1. Pre-procesamiento

La etapa de pre-procesamiento comienza con la aplicación de filtros clásicos con el objetivo de eliminar fuentes de ruidos indeseadas, tales como, línea de potencia, contracciones involuntarias musculares, ruido térmico, entre otros artefactos. Para eliminar estas componentes presentes en las señales ECG, se emplean diferentes tipos de filtrados, entre ellos, el filtro notch para eliminar línea base de potencia, filtro de respuesta al impulso finita (FIR) para remover componente de baja frecuencia por debajo de  $3Hz$  y alta frecuencia por encima de  $100Hz$ . Finalmente, los filtros adaptativos son usados para remover ruido estacionario, entre ellos el filtro Wiener y el filtro de Kalman producen una respuesta efectiva sin distorsionar la señal. Por otro lado, permite la supresión de ruido aplicando una descomposición tiempo-frecuencia. Luego de la descomposición la señal es reconstruida con sus componentes libres de ruido. Para garantizar el rendimiento de la transformada Wavelet se debe tener en cuenta aspectos como, elección de la función base, cálculo de coeficientes para determinar aquellos que representan ruido, así como los niveles de descomposición apropiados [31].

### 2.3.2. Algoritmo de Pan-tompkin

El algoritmo de detección de Pan-Tompkins es el método más robusto y confiable para la detección de complejos QRS, capaz de detectar los QRS en condiciones de mucho ruido o de señal muy alterada por causas fisiológicas. Involucra una serie de filtros y operadores de tipo paso bajo, paso alto, derivador, elevación al cuadrado, integrador, umbralización adaptativa y

reglas de decisión [8]. Las etapas del algoritmo se distribuyen de la siguiente manera:

Filtro paso banda: la señal ECG donde se pretende detectar los complejos QRS debe ser filtrada para obtener la banda donde se pueden encontrar los QRS y eliminar otro tipo de señales que interfieran. Primero se realiza un filtro paso bajo que elimina el ruido de alta frecuencia y después un filtro paso alto que elimina las ondas P y T [32].

Derivación: la señal filtrada, se deriva para detectar las pendientes pronunciadas características de los laterales de la forma de onda del complejo QRS [33].

Elevación al cuadrado: se eleva la señal ECG al cuadrado, lo que genera una señal positiva intensificando las altas frecuencias y atenuando las bajas, lo que distancia los complejos QRS de las ondas T de alta frecuencia [34].

Integración: la señal debe ser integrada mediante una ventana móvil. Es muy importante que el ancho de la ventana de integración sea aproximadamente igual al ancho del complejo QRS, ya que si la ventana se queda muy pequeña o grande la detección no es óptima. El valor de ancho de la ventana debe ser ajustado experimentalmente para cada ECG [35].

Reglas de decisión: si en la señal integrada se detecta un cambio de positivo a negativo, se determina un pico de energía. Luego, el algoritmo, usando umbrales adaptativos determinara si este pico de energía corresponde verdaderamente a un complejo QRS o debe ser considerado ruido. Los umbrales adaptativos o estimadores calculan durante el proceso de detección los valores promedios asociados a la amplitud de los picos del complejo QRS y del nivel de ruido. En función de estos dos valores, que se van actualizando con cada nueva detección, el algoritmo decide si un pico es ruido o es un QRS. Además, el algoritmo dispone de un sistema de búsqueda hacia atrás (en inglés, searchback) que permite la búsqueda hacia atrás de un pico R cuando no ha sido detectado durante un cierto periodo de tiempo después del pico R anterior [36].

### **2.3.3. Análisis espectral**

El análisis espectral es usado para determinar componentes presentes en una señal que no son perceptibles en el dominio del tiempo, es decir permite separar las diferentes componentes espectrales que conforman una señal para posteriormente poder analizarla en el dominio de la frecuencia [37]. Dentro de las técnicas usadas para analizar el espectro de una señal, se

encuentran las transformada de fourier(FT) y la transformada de wavelet(TW), entre otras [38].

#### 2.3.4. Transformada de Fourier

Esta herramienta matemática permite la transformación de señales del dominio del tiempo a la frecuencia y viceversa, lo cual hace que tenga diversas aplicaciones en ingeniería y física. Por ejemplo, permite analizar una señal en el dominio de la frecuencia para filtrar ruido o frecuencias indeseadas, como ocurre con las señales de audio o radiofrecuencias para las telecomunicaciones, es decir, calcula la contribución de cada valor de frecuencia a la formación de una señal continua. El concepto de esta transformada fue introducido por primera vez gracias a Joseph B. Fourier en 1811, todo esto a partir de las series de Fourier, en las cuales una función periódica se descompone en una suma de funciones trigonométricas[39]. La ecuación que describe la TF es:

$$F(\varpi) = \int_{-\infty}^{\infty} f(x)e^{-i\varpi x} dx \quad (2.3.1)$$

Donde se observa la integral que va desde  $-\infty$  hasta  $\infty$  de un producto entre  $e^{-i\varpi x}$ , que es una señal periódica y  $f(x)$  que es una señal de variable independiente continua [40]. Además, es importante tener en cuenta que  $i = \sqrt{-1}$  y  $e^{i\varpi x} = \cos(\varpi x) + i\sin(\varpi x)$ [41]

#### 2.3.5. Transformada Discreta de Fourier

Hace referencia al análisis de señales finitas y de tiempo discreto, utilizando la Transformada de Fourier[42]. Se define de la siguiente manera:

$$X(k) = \sum_{n=0}^{N-1} x(n)e^{-2\pi jkn/N} \quad (2.3.2)$$

De acuerdo a la ecuación, para cada  $k$  existen valores de 0 hasta  $N - 1$ . Además, cada componente espectral  $X(k)$  en los que se descompone una forma de onda  $x(n)$  con  $N$  muestras, puede hallarse calculando su ángulo de fase y amplitud, sumando los productos entre cada una de las muestras y la función sinusoidal compleja [42].

### 2.3.6. Transformada Rápida de Fourier

La transformada rápida de Fourier (FFT) es un algoritmo que permite calcular de manera rápida la Transformada Discreta de Fourier (DFT), eliminando gran parte de los cálculos iterativos que se hacen para la DFT. Fue introducida por J.W.Cooley y J.W Tukey en 1965, cuando abordaron por primera vez la implementación de un algoritmo para calcular series complejas [43].

La DFT que se desea calcular se divide en dos DFT menores a través de la FFT, teniendo en cuenta la paridad de los términos. A partir de la ecuación fundamental de la DFT 2.3.2, se aplica FFT y se obtiene una suma de dos DFT de  $N/2$  [44]. El resultado se observa en la siguiente ecuación:

$$X(k) = \sum_{r=0}^{N/2-1} x(2r)(W_{N/2})^{rk} + W_N^k \sum_{r=0}^{N/2-1} x(2r+1)(W_{N/2})^{rk} \quad (2.3.3)$$

Para  $r$  existen valores desde 0 hasta  $\frac{N}{2} - 1$  y  $W_N^2 = e^{\frac{-j2\pi}{N^2}} = e^{\frac{-j2\pi}{N/2}}$ .

### 2.3.7. Transformada de wavelet

La TW es usada para el análisis de señales no estacionarias y de rápida transitoriedad, asimismo, permite una representación de la señal en tiempo-frecuencia. La TW es usada, además para conocer el instante de tiempo asociado a cada contenido de frecuencia lo que le proporciona una ventaja frente la transformada de Fourier.

### 2.3.8. Transformada de wavelet continua

La transformada wavelet (CWT) permite la representación de señales en base a funciones bases seleccionadas de una familia de onduletas teniendo en cuenta la aplicación que se le quiere dar. La CWT usa un coeficiente de escala y de desplazamiento que permiten en análisis de resolución, esto es, mirar el contenido de frecuencia asociado a cada instante de tiempo. Con base a lo anterior, el rendimiento del método dependerá de la selección de la función madre y de los factores de escala y desplazamiento que estarán íntimamente relacionados con la naturaleza del problema. La CWT se calcula como sigue:

$$CWT = \frac{1}{\sqrt{a}} \int w\left(\frac{x-b}{a}\right) s(x) dx; \quad a, b \in \mathbb{R} \neq 0 \quad (2.3.4)$$

En donde  $a$  es el factor de escala y  $b$  el de desplazamiento y  $S(x)$  es la función a analizar

### 2.3.9. Transformada discreta de wavelet

La transformada discreta de wavelet (DWT) descompone una señal en varias escalas que representan diferentes bandas de frecuencia, filtros pasa altas y pasa bajas son usados bajos para la estimación de los coeficientes de detalle y aproximación respectivamente. La ecuación 2.3.5 representa la DWT con una función base  $\psi(t)$  y factor de escalado  $\phi(t)$ , el primer apartado corresponde a los coeficientes de detalle, mientras que el segundo los coeficientes de detalle. Finalmente la señal queda representada como la sumatoria de todos los componentes de la descomposición lo que permite la reconstrucción [45, 46].

$$f(t) = \sum_k \sum_j c_{j,k} \phi_{j,k}(t) + \sum_k \sum_j d_{j,k} \psi_{j,k}(t) \quad (2.3.5)$$

## 2.4. Medidas de dispersión

Las medidas de dispersión estadística permiten observar que tanta variación presenta una variable respecto a la tendencia de los datos definidos. La tendencia central de los datos es estimada principalmente por medidas como, mediana, media y moda. De esta forma las variables de estudio son caracterizadas en base a su rango de dispersión [47].

Dentro de las medidas de dispersión se encuentra, rango, rango intercuartil, desviación estándar, varianza y media de la desviación absoluta, entre otras.

### 2.4.1. Rango

Esta medida representa la diferencia que hay entre un valor máximo y mínimo de una muestra estadística. Se define como:

$$R = X_{max} - X_{min}$$

(2.4.1)

Donde  $R$  es el rango,  $X$  la variable,  $X_{\min}$  el valor mínimo de la muestra y  $X_{\max}$  el valor máximo de la muestra.

### 2.4.2. Rango intercuartil

Representa la diferencia entre el tercer y primer cuartil de un conjunto de datos, de esta manera divide la distribución de datos en cuatro partes iguales. Se define de la siguiente manera:

$$IQR = Q_3 - Q_1 \quad (2.4.2)$$

Donde  $IQR$  es el rango intercuartil,  $Q_3$  es el tercer cuartil y  $Q_1$  el primer cuartil.

### 2.4.3. Desviación estándar

Hace referencia al valor medio de las desviaciones o valores alejados de la media, de un conjunto de datos. Su fórmula es la siguiente:

$$\sigma = \sqrt{\frac{\sum_1^n (x_i - \bar{X})^2}{n}} \quad (2.4.3)$$

Donde  $x$  es la variable,  $x_i$  es cada valor de la variable según el conjunto de datos dispersos,  $\bar{X}$  es la media de la variable y  $n$  el número de observaciones.

### 2.4.4. Varianza

Esta medida de dispersión, representa la variación de un conjunto de datos respecto a la media, matemáticamente se presenta como la desviación estándar elevada al cuadrado. Su fórmula es la siguiente:

$$\sigma = \frac{\sum_1^n (x_i - \bar{X})^2}{n} \quad (2.4.4)$$



Donde  $x$  es la variable,  $x_i$  es cada valor que toma la variable según el conjunto de datos,  $\bar{X}$  es la media de la variable y  $n$  el número de observaciones.

#### 2.4.5. Media de la desviación absoluta

Representa la desviación promedio del conjunto de datos, es decir la distancia entre cada valor y el promedio. Nos permite conocer que tan dispersos están los valores en la muestra o población. Se define matemáticamente de la siguiente manera:

$$MAD = \frac{\sum_{i=1}^g |x_i - \bar{X}| f_i}{n} \quad (2.4.5)$$

Donde  $g$  son las clases de distribución de frecuencia o número intervalos,  $X_i$  es marca de clase de cada intervalo,  $f_i$  número de observaciones clasificadas en cada clase.

### 2.5. Métodos de inteligencia artificial

La inteligencia artificial (IA), busca estudiar y analizar el comportamiento humano en los ámbitos de la comprensión, percepción, resolución de problemas y de la toma de decisiones con el fin de poder reproducirlos con la ayuda de un computador. De esta manera, las aplicaciones de la IA se sitúan principalmente en la simulación de actividades intelectuales del hombre. Es decir, imitar por medio de máquinas digitales, tantas actividades mentales de toma de decisiones como sea posible [48].

De igual manera, la IA es una disciplina académica directamente relacionada con la “teoría de la computación”, la cual tiene como principal objetivo emular algunas de las facultades intelectuales humanas en los sistemas artificiales. Ahora bien, para la implementación de un sistema de IA, normalmente se requiere la utilización de herramientas de diferentes campos para extraer características, tales como cálculo numérico, estadística, informática, procesado de señales, neurociencia, entre otras. Sin embargo, pese a que la IA se considera una rama estricta de la informática teórica, es una disciplina en la que participan diferentes profesionales, de manera activa, tales como científicos, técnicos, matemáticos, entre otros [49, 50]. Particularmente, la IA supone un importante esfuerzo por emular el proceso de toma de decisiones inteligentes de los seres humanos en términos de procesos de información. Asimismo, se trata de un área extraordinariamente vital que experimenta un crecimiento

acelerado abarcando otras áreas lo que la hace una herramienta multidisciplinaria.

### **2.5.1. Machine learning**

Los algoritmos de aprendizaje automático permiten aprender un modelo a partir de los datos y utilizarlo para resolver tareas prácticas de la minería de datos, la cual es utilizada para extraer información a partir de bases de datos, generando patrones nuevos, útiles y comprensibles que son usados como predictores. El objetivo es encontrar patrones fuertes, aquellos que pueden hacer predicciones con alto grado de exactitud en diferentes campos de la vida cotidiana [51].

Asimismo, se trata de una disciplina científica del ámbito de la IA que crea sistemas que aprendan automáticamente. Con aprender, dentro de este contexto, se hace referencia a algoritmos que automaticen sistemas complejos en una densidad datos [52].

Deep learning, o modelos de aprendizaje profundo, es la parte del aprendizaje automático que, por medio de algoritmos de alto nivel, imita la red neuronal del cerebro humano, se componen de muchos modelos lineales simples (nodos) dispuestos en serie (cada serie se denomina capas), que procesan la información dada en múltiples etapas sucesivas para así tener en cuenta interacciones complejas entre los datos observados, La principal aplicación de los algoritmos de Deep Learning son las tareas de clasificación, en especial, reconocimiento de imágenes [53].

### **2.5.2. Big data**

Es una referencia a aquellos sistemas de información que manejan conjuntos de datos de gran volumen, de alta velocidad, de veracidad, de valor y de gran variedad de recursos, que demandan formas rentables e innovadoras de procesamiento de la información para mejorar la comprensión y la toma de decisiones [54]. Asimismo, los “datos masivos”, tienen la capacidad de aprovechar la información de los usuarios a nivel mundial de forma novedosa por medio de los servicios de la nube, con el fin de obtener percepciones útiles que resulten significativas en lo que se esté necesitando en los diferentes campos del conocimiento [55].

Por otro lado, el big data puede ser utilizado para las empresas y organizaciones que buscan obtener la mayor cantidad de información para obtener alguna ventaja sobre la competencia. Por lo anterior, los datos reflejados dentro del big data deben manejarse con precisión en cuanto a su análisis. Asimismo, dentro del uso biomédico, se tienen en cuenta los métodos y

las ventajas de los macrodatos y su funcionabilidad dentro de la extracción de características [56].

El big data está siendo aplicado actualmente en el almacenamiento de gran cantidad de registros médicos electrónicos, como imágenes diagnósticas, electrocardiogramas (ECG), electroencefalogramas (EEG), electromiogramas (EMG), historias clínicas, entre otros. Esto no solo permite la creación de bases de datos masivas sino también la aplicación de diferentes métodos de análisis para los datos almacenados. De esta forma, en conjunto con IA es posible desarrollar nuevas tecnologías para la prevención, diagnóstico y tratamiento de diferentes enfermedades [57]. Por ejemplo, un médico especialista, desde una ubicación remota puede ver los registros clínicos de cualquier paciente a través de dispositivos inteligentes, así dar un diagnóstico oportuno y realizar los procesos necesarios para el tratamiento del paciente, en el caso de requerirse [58, 59].

## 2.6. Análisis de Componentes Principales

Cuando se tiene una amplia cantidad de datos que pertenecen a una población estadística en estudio, es necesario aplicar métodos que permitan la reducción de valores redundantes en el conjunto de datos, para esto se utiliza el análisis de componentes principales (en inglés PCA), ya que permite representar un conjunto de datos correlacionados a través de nuevas variables con menor correlación. También, es necesario tener en cuenta que entre mayor varianza exista entre el conjunto de datos, mayor información se tiene, por el contrario, si la varianza es baja quiere decir que existen muchas variables redundantes o con valores similares [60].

Entre a los conceptos matemáticos implementados en PCA se encuentran principalmente, eigenvectors y eigenvalues.

### 2.6.1. Eigenvectors

Hace referencia a los vectores de una matriz, los cuales al multiplicarlos por la misma se obtiene un vector resultante que puede ser el mismo vector o un múltiplo entero del vector original[61]. La operación es similar a la siguiente:

$$\begin{pmatrix} 2 & 3 \\ 2 & 1 \end{pmatrix} \times \begin{pmatrix} 3 \\ 2 \end{pmatrix} = \begin{pmatrix} 12 \\ 8 \end{pmatrix} = 4 \times \begin{pmatrix} 3 \\ 2 \end{pmatrix}$$

### 2.6.2. Eigenvalues

Del producto entre una matriz y sus eigenvectors se obtiene el mismo vector multiplicado por un número, el cual tiene un valor que se conoce como eigenvalue. Al utilizar PCA cada eigenvector representa una componente y el orden jerárquico para organizar cada componente se determina por cada eigenvalue [61].

### 2.6.3. Cálculo de las componentes principales

Este cálculo se hace a partir de un conjunto de variables que van de  $X_1$  hasta  $X_p$ , donde  $p$  es la dimensión del espacio muestral, así se pretende hallar un conjunto de datos que va de  $Z_1$  hasta  $Z_p$ , con menor correlación y ordenado de mayor a menor en cuanto a la varianza de cada uno de los datos [62]. Cada componente  $Z_i$  del nuevo conjunto de datos es una combinación lineal normalizada del conjunto de datos original[61]. Por tanto, para la primera componente del nuevo conjunto de datos, se tiene que:

$$Z_1 = \phi_{11}X_1 + \phi_{21}X_2 + \dots + \phi_{p1}X_p \quad (2.6.1)$$

Cada  $\phi_{p1}$  se conoce como *loading*. El conjunto de *loadings* permite conocer el tipo de información recogida por cada componente principal.

Además, ya que cada componente del nuevo conjunto de datos es una combinación lineal normalizada, se tiene que:

$$\sum_{j=1}^p \phi_{j1}^2 = 1 \quad (2.6.2)$$

## Capítulo 3

# Metodología

Las siguientes etapas y actividades han sido definidas para la ejecución del proyecto:

Fase 1: Definición de las especificaciones generales del sistema.

Definición de los requisitos y especificaciones del sistema. Con base en la descripción del problema y el estado de la técnica, se determina el requisito general para el sistema de detección automática de pacientes con arritmias auriculares.

Fase 2: Desarrollo del sistema.

- Base de datos Los registros ECG son obtenidos de bases de datos con AF, AFL y ritmo sinusal normal. Las señales son muestreadas a una frecuencia de muestreo de 1 KHz con 12 canales de derivación. Las señales de arritmia auricular corresponden a un estudio llevado en España por un investigador local en el hospital universitario de Valencia, mientras que las señales de ritmo sinusal normal han sido tomadas de la colección PTB de physionet.
- Detección de los complejos QRS Diferentes técnicas de detección de complejos QRS serán evaluadas basadas en el análisis de la forma de onda y herramientas estadísticas en el dominio del tiempo y la frecuencia. Será elegida la que presente el mejor rendimiento detectando complejos QRS tanto en las señales de ritmo sinusal normal como las de arritmia.
- Medidas de dispersión para cuantificar la variabilidad de la frecuencia cardiaca

Diferentes medidas de dispersión serán usadas una vez que sea calculada la distancia entre los intervalos R-R, las medidas de dispersión deben resaltar la diferencia que existe entre

la variabilidad de la frecuencia cardiaca para señales con arritmia auricular y señales de ritmo sinusal normal. Serán elegidas las medidas de dispersión que describan mejor el comportamiento anterior, esto es, que muestren una alta dispersión para las señales con arritmia auricular y una más baja para las señales de ritmo sinusal normal.

- Características espectrales

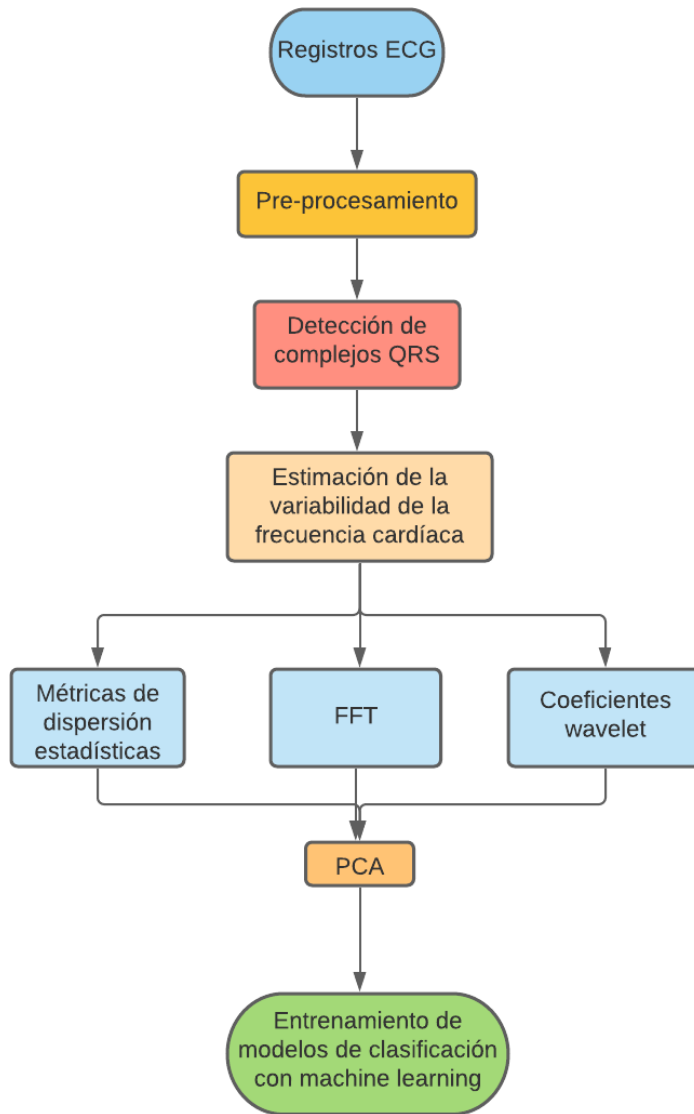
Una vez seleccionadas las medidas de dispersión, se eligen características espectrales que permitan resaltar la diferencia que existe en cuanto al espectro de las señales de ritmo sinusal normal y arritmia cardiaca. Basados en la literatura son seleccionadas la densidad espectral de potencia, pues se sabe que las señales con arritmia cardíaca tienen más armónicos que las señales de ritmo sinusal normal, además se selecciona la frecuencia fundamental ya que se sabe que en ritmo sinusal normal la frecuencia fundamental se encuentra siempre en el componente relacionado con los complejos QRS y este tiende a ser constante, mientras que en las señales con arritmia auricular esta frecuencia fundamental puede estar desplazada y no ser tan dominante. Además, se añaden los coeficientes wavelet con tres niveles de descomposición usando la wavelet madre db2, con el fin de resaltar la diferencia que existe en la actividad ventricular de cada clase.

- Aplicación de las herramientas de inteligencia artificial

Serán probados todos los métodos de machine learning de los que dispone el toolbox de inteligencia artificial del lenguaje MATLAB, para mirar el rendimiento de algunas características los experimentos serán llevados a cabo en el siguiente orden:

1. Se entrenan y prueban los clasificadores únicamente con las medidas de dispersión, la densidad espectral de potencia y la frecuencia fundamental.
2. Se entrenan y prueban los clasificadores únicamente con las medidas de dispersión, y los coeficientes wavelet.
3. Se entrenan y prueban los clasificadores con todas las medidas.
4. Se entrenan y prueban los clasificadores usando las primeras 5 características que arroja el análisis de componentes principales.

A continuación, se presenta gráficamente la estructura del software a desarrollar:



**Figura 3.0.1:** Diagrama de bloques para el desarrollo del software.

### 3.1. Pre-procesamiento

Inicialmente, se creó una matriz con todas las señales de fibrilación auricular y ritmo sinusal normal. Para poder analizar cada muestra de las bases de datos mencionadas anteriormente, a través de las técnicas adecuadas, es necesario en primer lugar filtrar el ruido existente en cada una de ellas. En este caso, se aplicó un filtro butterworth de orden 3 y frecuencia de corte de  $1Hz$ , utilizando el toolbox para procesamiento de señales en MATLAB (Signal Processing Toolbox). De esta forma, se logró corregir la línea base y eliminar el ruido generado por la

actividad muscular.

## 3.2. Detección de complejos QRS usando el algoritmo de pan tompkins

Para la detección de los complejos QRS de cada señal se utilizó el algoritmo de Pan-Tompkins. Inicialmente un filtro derivativo es aplicado a la señal para resaltar los complejos QRS y eliminar componentes de baja frecuencia asociados a las ondas P y T. Posteriormente, un filtro de cuadratura es aplicado a la señal para amplificar las grandes diferencias resultante de los complejos QRS, así como, suprimir componentes remanentes relacionados a las ondas P y T eliminadas en la etapa anterior. Después de aplicar el filtro de cuadratura, la operación de integración es llevada a cabo usando una ventana móvil con el objetivo de alisar la señal cuando varios picos se encuentran presentes en un mismo latido. Por último, está la etapa de Regla de Decisión, la cual a su vez se divide en diferentes etapas.

### 3.2.1. Regla de Decisión

En esta parte del algoritmo, las etapas anteriores han producido una forma de onda aproximadamente en forma de pulso en la salida de la etapa de integración. La determinación de si este pulso corresponde a un complejo QRS se realiza con una operación de umbral adaptativo y otras reglas de decisión, las cuales son las siguientes:

### 3.2.2. Marca Fiducial

En esta etapa, la forma de onda se procesa primero para producir un conjunto de muestras unitarias ponderadas en la ubicación de los máximos de la señal después de aplicar la etapa de integración. Esto se hace para localizar el complejo QRS en un solo instante de tiempo. La ponderación  $w[k]$  es el valor máximo.

### 3.2.3. Umbral

Al analizar la amplitud de la salida de la etapa de integración, el algoritmo utiliza dos valores de umbral  $THR_{SIG}$  y  $THR_{NOISE}$ , adecuadamente inicializados durante una breve fase de entrenamiento de 2 segundos que se adaptan continuamente a los cambios en la calidad de



la señal ECG. La primera pasada por  $y[n]$  utiliza estos umbrales para clasificar cada muestra distinta de cero ( $CURRENTPEAK$ ) como señal o ruido:

Si  $CURRENTPEAK > THR_{SIG}$ , ese lugar se identifica como un candidato a complejo QRS y se actualiza el nivel de señal ( $SIG_{LEV}$ ):

$$SIG_{LEV} = 0,125CURRENTPEAK + 0,875SIG_{LEV} \quad (3.2.1)$$

Si  $THR_{NOISE} < CURRENTPEAK < THR_{SIG}$ , ese lugar se identifica como un pico de ruido y se actualiza el nivel de ruido ( $NOISE_{LEV}$ ):

$$NOISE_{LEV} = 0,125CURRENTPEAK + 0,875NOISE_{LEV} \quad (3.2.2)$$

Basándose en las nuevas estimaciones de los niveles de señal y ruido ( $SIG_{LEV}$  y  $NOISE_{LEV}$ , respectivamente) en ese punto del análisis de la señal ECG, los umbrales son ajustados de la siguiente manera:

$$THR_{SIG} = NOISE_{LEV} + 0,25 \quad (3.2.3)$$

$$THR_{NOISE} = 0,5(THR_{SIG}) \quad (3.2.4)$$

Estos ajustes reducen el umbral gradualmente en los segmentos de señal que se consideran de menor calidad.

#### 3.2.4. Búsqueda de los complejos QRS faltantes

Basado en la anterior etapa de umbralización, si  $CURRENTPEAK < THR_{SIG}$ , se considera que el pico no es un complejo QRS. Sin embargo, si ha transcurrido un periodo excesivamente largo sin un pico por encima del umbral, el algoritmo asumirá que se ha perdido un QRS y realizará una búsqueda de vuelta. Esto limita el número de falsos negativos. El tiempo mínimo utilizado para desencadenar una búsqueda de vuelta es 1,66 veces el período de tiempo entre picos R actuales (denominado intervalo RR). Este valor tiene un origen fisiológico, debido

a que el valor del tiempo entre latidos adyacentes no puede cambiar más rápido que eso. Se supone que el complejo QRS perdido se produce en la ubicación del pico más alto del intervalo que se encuentra entre  $THR_{SIG}$  y  $THR_{NOISE}$ . En este algoritmo, se almacenan dos intervalos RR promedio, el primer intervalo RR se calcula como un promedio de las últimas ocho ubicaciones de QRS para adaptarse a los cambios en la frecuencia cardíaca y el segundo intervalo RR promedio es la media de los intervalos RR más regulares. El umbral se reduce si la frecuencia cardíaca no es regular para mejorar la detección.

### 3.2.5. Eliminación de detecciones múltiples dentro del periodo refractorio

Es imposible que se produzca un complejo QRS legítimo si se encuentra dentro de los 200 ms siguientes a otro detectado previamente. Esta restricción es fisiológica debido al periodo refractorio durante el cual la despolarización ventricular no puede producirse a pesar de un estímulo. A medida que se generan los complejos QRS candidatos, el algoritmo elimina estos eventos físicamente imposibles, reduciendo así los falsos positivos.

### 3.2.6. Discriminación de la onda T

Por último, si un candidato a QRS se produce después del periodo refractorio de  $200ms$  pero dentro de los  $360ms$  del QRS anterior, el algoritmo determinará si se trata acerca de un auténtico complejo QRS del siguiente latido o de una onda T anormalmente prominente. Esa decisión se toma basado en la pendiente media de la forma de onda en esa posición. Una pendiente inferior a la mitad de la del complejo QRS anterior es coherente con el comportamiento de cambio más lento de una onda T, de lo contrario, se convierte en un complejo QRS detectado. Esta etapa del algoritmo, también comprueba si el pico que se produce, cuya latencia es inferior a  $360ms$ , tiene también una latencia inferior a  $0,5 * mean_{RR}$ , de ser así, se cuenta como ruido.

En la etapa final, la salida de las ondas R detectadas en la señal suavizada, se analiza y se comprueba dos veces gracias a la señal de salida del filtro paso-banda, para mejorar la detección y encontrar el índice original de las ondas R en la señal de ECG sin procesar.

### **3.3. Estimación de la variabilidad de la frecuencia cardíaca usando medidas de dispersión**

A partir de los resultados obtenidos del algoritmo de Pan- Tompkins se obtienen los intervalos de las posiciones donde se encuentran ubicados los picos R. En base a lo anterior, la variabilidad de la frecuencia cardíaca es estimada mediante la diferencia de todos los picos R detectados, generando así los intervalos RR. Para conocer la variación entre los intervalos RR de cada señal, varias medidas de dispersión son estimadas, entre estas: desviación estándar, rango, rango intercuartil y desviación de la media absoluta. Esto permitió identificar las variaciones en frecuencia cardíaca para ritmo sinusal normal y arritmia auricular. Basado en la literatura se espera obtener mayores valores de dispersión para las señales con arritmias auriculares debido a la variabilidad de los intervalos RR, así como valores más bajo de dispersión para las señales de ritmo sinusal normal, ya que en esta condición los intervalos RR son constantes.

También, se usó el análisis espectral para estimar la frecuencia dominante y la densidad espectral de potencia de cada señal. Se espera que esta característica permita distinguir entre las dos clases, ritmo sinusal normal y arritmia, ya que el espectro de las señales con arritmia presenta mayor cantidad de armónicos que el espectro de las señales de ritmo sinusal normal, así como una frecuencia dominante desplazada para las señales con arritmia debido a la variabilidad de los complejos QRS.

Finalmente, la descomposición wavelet es utilizada, usando la wavelet madre daubechies 2 con 3 niveles de descomposición, para poder apreciar los coeficientes de detalles relacionados a los complejos QRS. Con base en la variabilidad de la actividad ventricular y la falta de sincronización de la actividad auriculo-ventricular en las señales de arritmia, se espera una diferencia entre los coeficientes asociados a cada ritmo que permita distinguir entre las clases.

### **3.4. Creación del dataset**

En esta etapa se creó una matriz para agrupar las características. Esta matriz distribuye las características en 8 columnas principales, las cuales se encuentran agrupadas en el siguiente: medidas de dispersión asociadas a la variabilidad de la frecuencia cardíaca, desviación estándar

(*STD*), rango intercuartil (*RQI*), rango (*RNG*), media (*MED*), media de la desviación absoluta (*MAD*). Características espectrales, densidad espectral de potencia (*PSD*) y frecuencia dominante ( $f_0$ ). Finalmente, la matriz es completada añadiendo los coeficientes de detalle wavelet (*CDW*).

Para reducir la dimensionalidad del conjunto de datos, se utiliza el PCA y como resultado final se obtiene una matriz con los eigenvalores del conjunto de características. En orden de hacer un análisis comparativo de las técnicas, los experimentos son llevados a cabo en los siguientes escenarios: Toda la matriz de características sin (*CDW*), toda la matriz de características sin (*PSD*) y ( $f_0$ ), la matriz de características únicamente con las medidas dispersión y de tendencia central, y finalmente usando la matriz de eigenvalores arrojada por el PCA tomando las primeras 5 valores principales.

### **3.5. Aplicación de los métodos de inteligencia artificial**

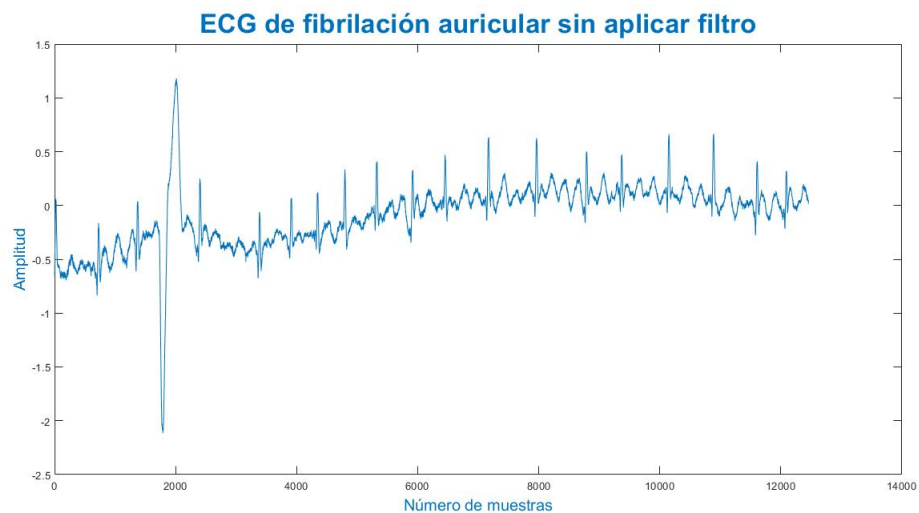
Finalmente, se introdujo la matriz de características en el aplicativo Classification Learner de MATLAB, se entrenaron todos los modelos de clasificación disponibles en la app, para mirar el rendimiento de los métodos se analiza los resultados arrojados por la app de MATLAB, estos son, matriz de confusión, gráfico de dispersión, diagrama de coordenadas paralelas, entre otros. Los experimentos fueron realizados con PCA y sin PCA para mirar el desempeño de cada característica.

# Capítulo 4

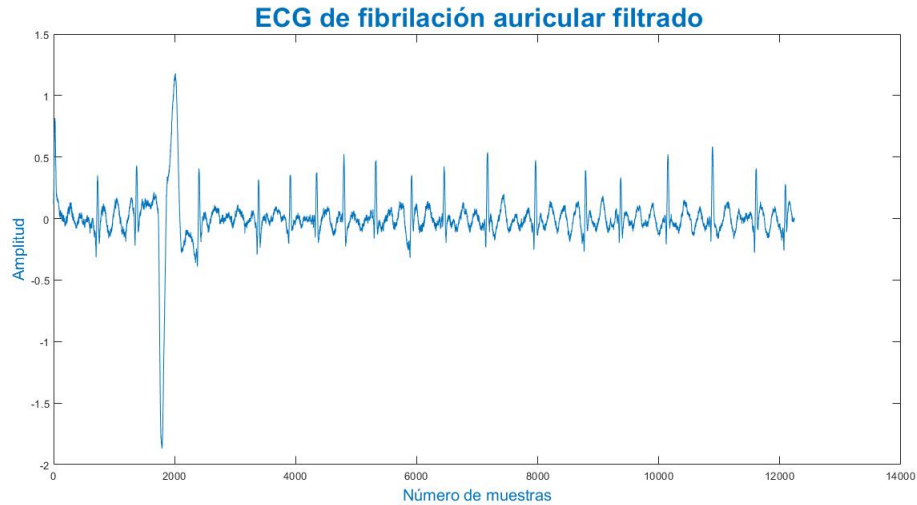
## Resultados

### 4.1. Pre-procesamiento

En la etapa de pre-procesamiento, un filtro butterworth paso- alto de orden 3 y frecuencia de corte de  $1Hz$  es aplicado a cada uno de los 99 ECG obtenidos de las bases de datos, 49 correspondientes a arritmias auriculares y 50 correspondientes a ritmo sinusal normal, se logra eliminar el ruido generado por la actividad muscular y corregir la línea base en cada uno de ellos. En la figura 4.1.1, se observa un ECG de arritmia auricular sin ser filtrado y en la figura 4.1.2 se observa el resultado luego de aplicar el filtro.



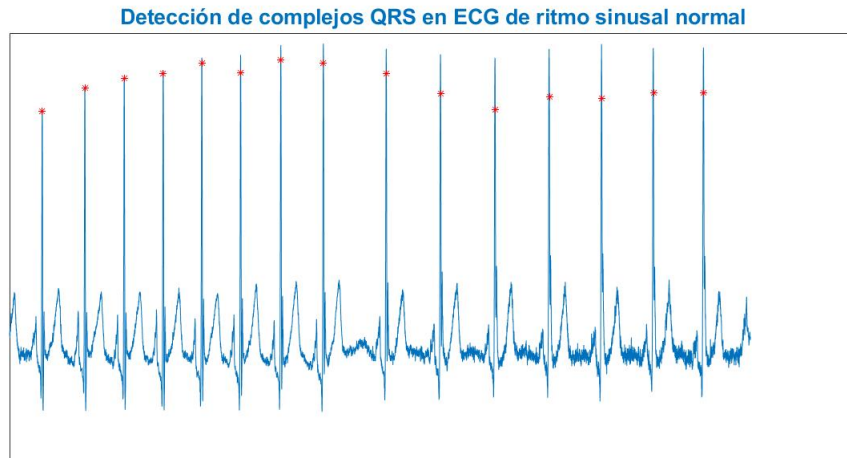
**Figura 4.1.1:** ECG de fibrilación auricular contenido en el archivo x003.mat, sin corrección de línea base y con ruido producido por actividad muscular.



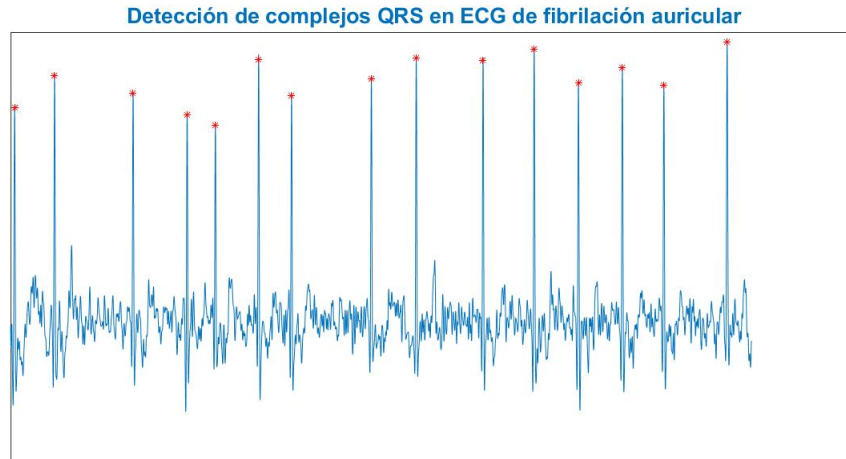
**Figura 4.1.2:** ECG de fibrilación auricular contenido en el archivo x003.mat, luego de aplicar filtro butterworth.

## 4.2. Detección de complejos QRS

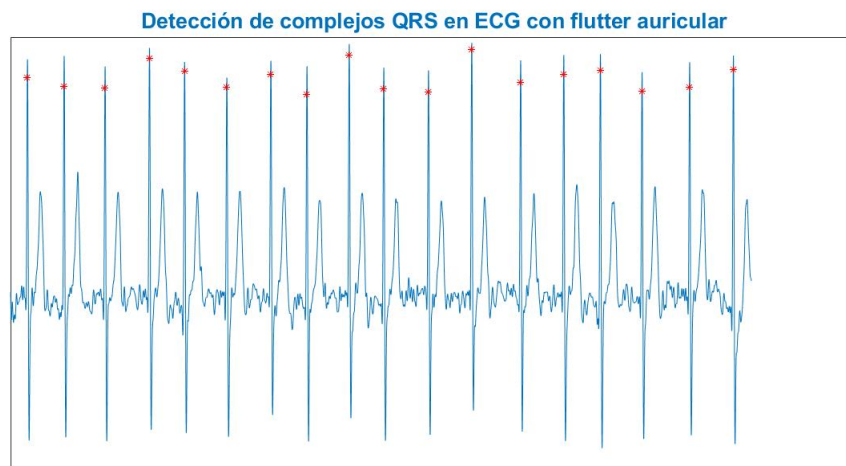
La detección de los complejos QRS es llevada a cabo mediante el algoritmo de Pan- Tompkins, los resultados obtenidos se observan en la figura 4.2.1, figura 4.2.2 y figura 4.2.3, en las cuales se destacan los complejos QRS para cada tipo señal, una corresponde a ritmo sinusal normal y dos de arritmias auriculares. En cada figura los puntos marcados en color rojo son los picos R de cada complejo QRS detectado.



**Figura 4.2.1:** Detección de complejos QRS en ECG de ritmo sinusal normal a través del algoritmo de Pan-Tompkins.



**Figura 4.2.2:** Detección de complejos QRS en ECG de fibrilación auricular a través del algoritmo de Pan-Tompkins.

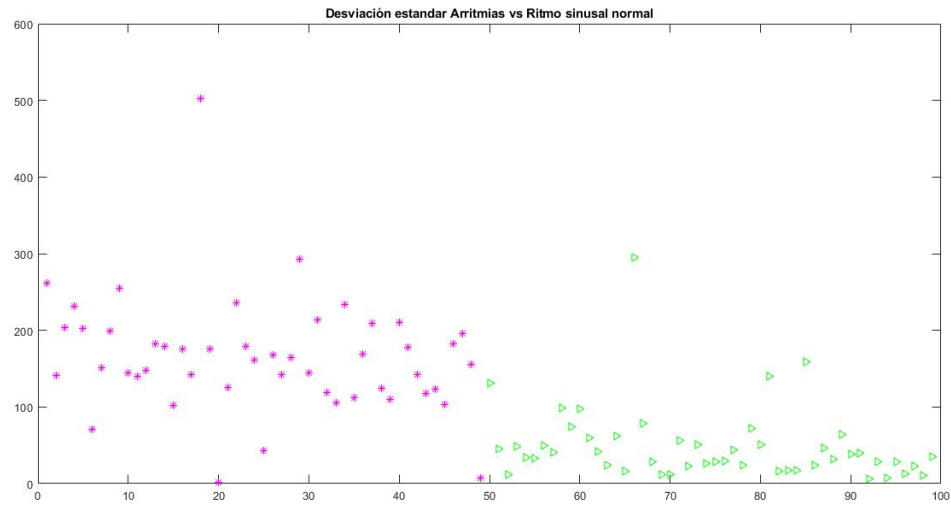


**Figura 4.2.3:** Detección de complejos QRS en ECG de flutter auricular a través del algoritmo de Pan-Tompkins.

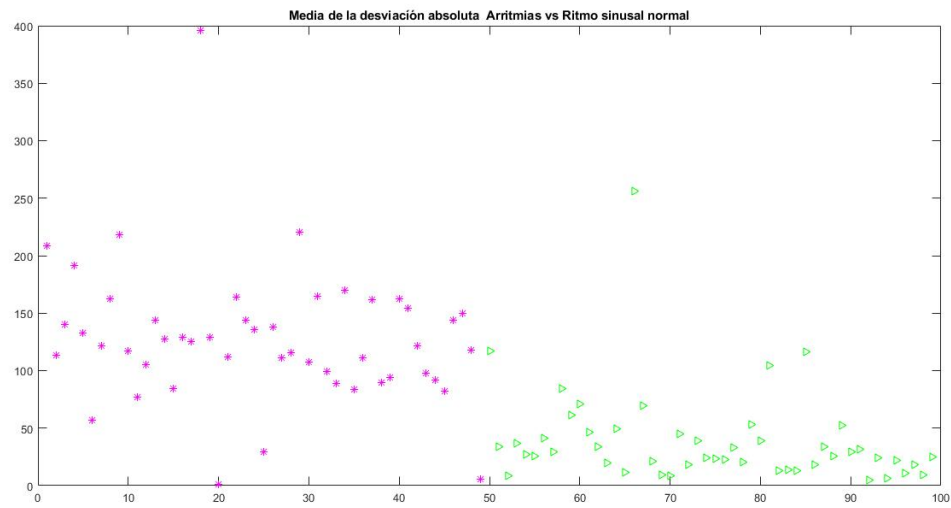
### 4.3. Extracción de características

De acuerdo a las medidas de dispersión utilizadas para extraer características, se observa que la desviación estándar, la media de la desviación absoluta, el rango y el rango intercuartil tienen valores mayores para los intervalos R-R en los ECG de arritmias auriculares, en contraste con los ECG de ritmo sinusal normal como se observa desde la figura 4.3.1 hasta la figura 4.3.4. Asimismo, se observa en el eje X de cada figura el número de señales y en el eje

Y los valores de cada medida de dispersión para cada una de ellas. También, se clasifican las señales utilizando el color verde para ritmo sinusal normal y fucsia para arritmias auriculares.

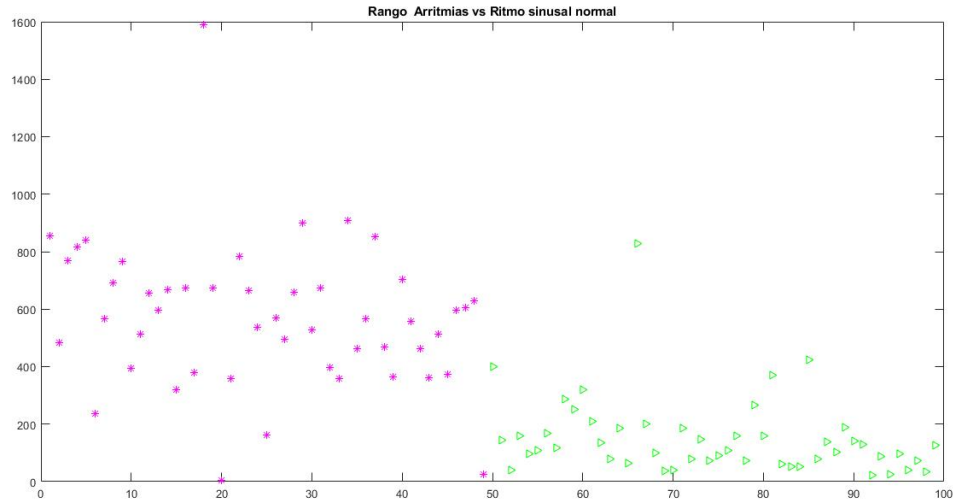


**Figura 4.3.1:** Gráfico comparativo para la desviación estandar entre los ECG.

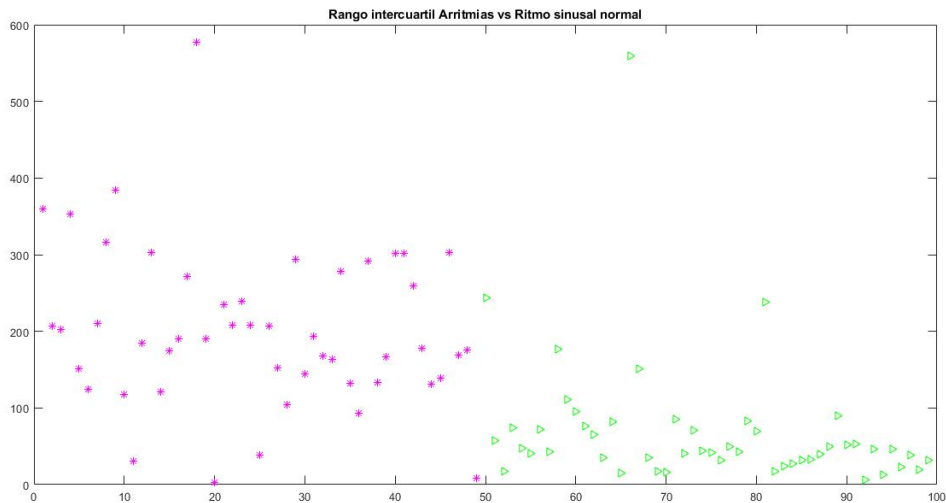


**Figura 4.3.2:** Gráfico comparativo para la media de la desviación absoluta entre los ECG.



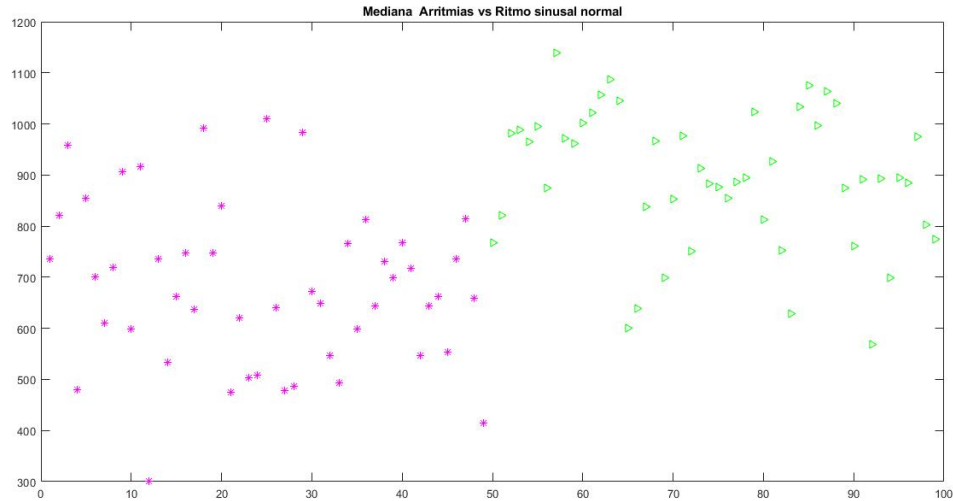


**Figura 4.3.3:** Gráfico comparativo para el rango entre los ECG.



**Figura 4.3.4:** Gráfico comparativo para el rango intercuartil entre los ECG.

Asimismo, se observó que los valores para la mediana no varían demasiado en las señales de ritmo sinusal normal en contraste con las arritmias auriculares. En la figura 4.4.1 se muestra la comparación entre ambas clases de señales.



**Figura 4.3.5:** Gráfico comparativo para la mediana entre los ECG.

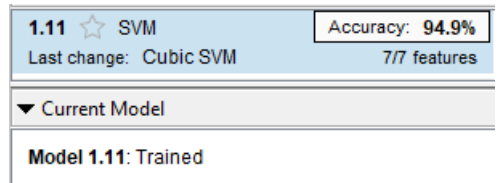
De acuerdo a la gráfica de la figura 4.4.1, en el eje X se observa el número de señales y en el eje Y se observan los valores correspondientes a la mediana para cada una de ellas. Asimismo, se clasifican las señales utilizando el color verde para ritmo sinusal normal y fucsia para arritmias auriculares.

## 4.4. Inteligencia artificial

A partir de las características extraídas a través de las medidas de dispersión, FFT y DWT, se obtuvo diferentes resultados en el aplicativo Classification Learner de MATLAB, los cuales se dividen en cuatro casos, en el primero se presentan los resultados utilizando solo medidas estadísticas y FFT, en el segundo se utiliza las medidas estadísticas y wavelet, en el tercero se incluyen todos los métodos de extracción de características y en el último caso se incluyen solo 5 eigenvalores obtenidos con PCA.

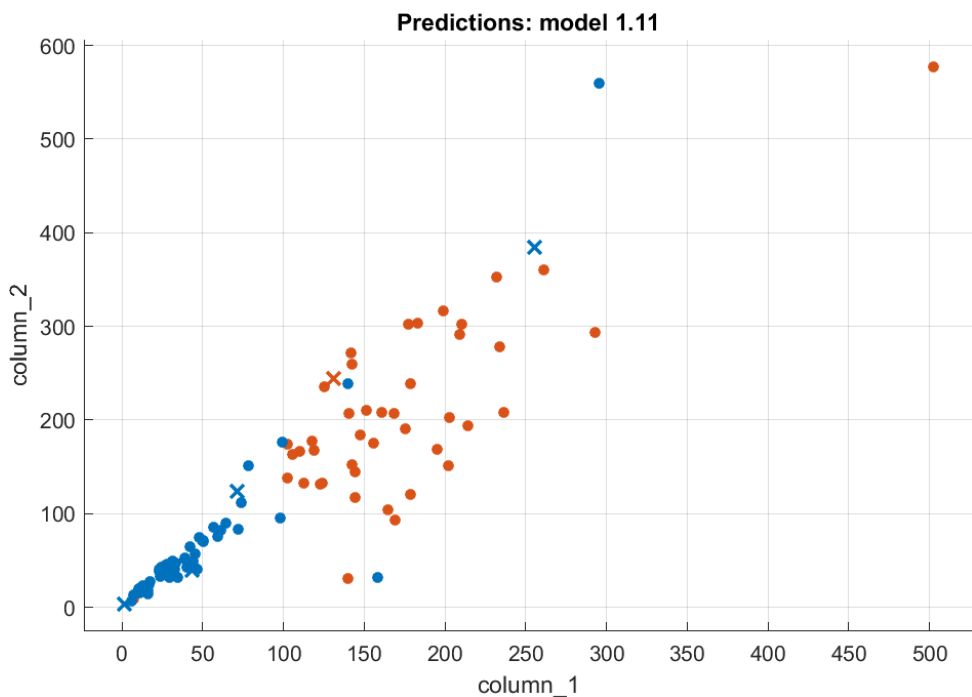
### 4.4.1. Caso 1

En este caso se utilizan únicamente las medidas de dispersión, las de tendencia central, la densidad espectral y la frecuencia fundamental. En la figura 4.4.1 se observa la mejor precisión obtenida entre los modelos de clasificación disponibles en el aplicativo Classification Learner.



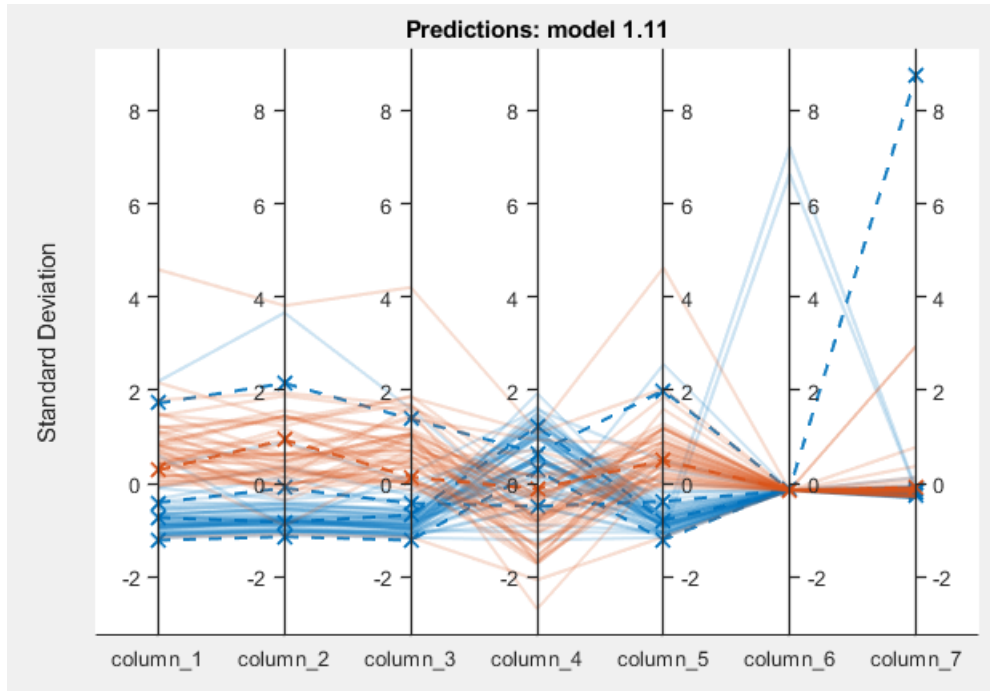
**Figura 4.4.1:** Resultado obtenido para el modelo de clasificación Cubic SVM. Presenta una precisión de 95 %.

También, se presenta en los resultados la gráfica de dispersión, las coordenadas paralelas, la matriz de confusión y la región de convergencia para este caso. Obsérvese las siguientes figuras:



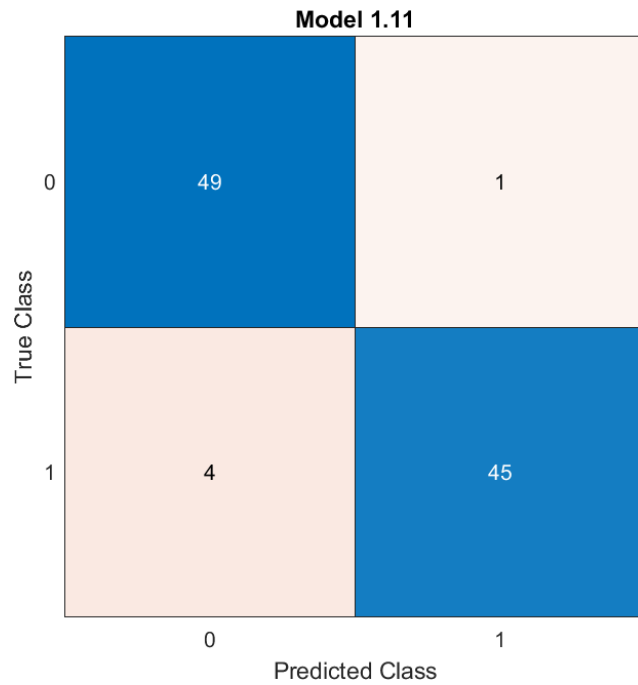
**Figura 4.4.2:** Gráfica de dispersión para el caso 1.

De acuerdo a la figura 4.4.2 se presentan dos características,  $column_1$  en el eje X es la desviación estandar y  $column_2$  en el eje Y representa el rango. Los puntos de color naranja representan las arritmias auriculares y los de color azul los ritmos sinusales normales. También, se destaca que los valores para las dos características presentadas estan menos dispersos en las señales con ritmo sinusal normal.



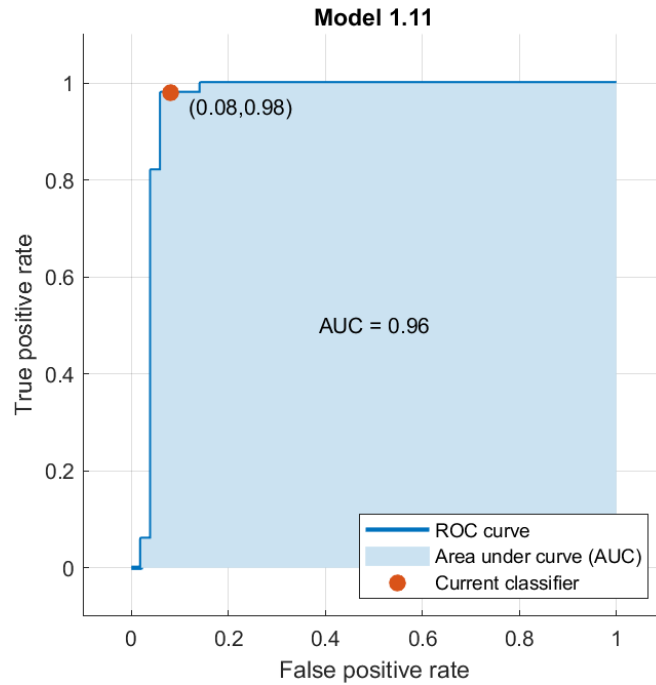
**Figura 4.4.3:** Coordenadas paralelas para el caso 1.

En figura 4.4.3 se presentan las coordenadas paralelas para las arritmias auriculares y los ritmos sinusales normales. Las líneas color naranja representan las arritmias auriculares, las de color azul los ritmos sinusales normales y desde  $column_1$  hasta  $column_7$  se muestran las 7 características pertenecientes al caso 1. Asimismo, se observa una menor relación entre los valores para cada una de las características de las señales de arritmias auriculares en comparación con las de ritmos sinusales normales.



**Figura 4.4.4:** Matriz de confusión para el caso 1.

De acuerdo a la matriz de confusión de la figura 4.4.4, se observa que 49 señales fueron correctamente clasificadas como ritmo sinusal normal y erróneamente 4 como arritmias auriculares. Asimismo, se clasificaron correctamente 45 señales como arritmias auriculares y erróneamente 1 como ritmo sinusal normal.

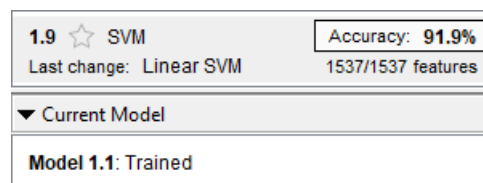


**Figura 4.4.5:** Región de convergencia para el caso 1

La figura 4.4.5 corresponde a la región de convergencia, en la cual se observa la curva ROC en color azul oscuro, el área bajo la curva (*AUC*) en color azul celeste, el punto naranja representa al clasificador, el eje X el ratio de falsos positivos (*FPR*) y el eje Y el ratio de verdaderos positivos (*TPR*). También, se observa que el 98 % de señales fueron correctamente clasificadas en la clase positiva frente a un 0,8 % que fueron clasificadas erróneamente en esta clase. También se observa un rendimiento del 96 % para el clasificador, valor descrito por *AUC*.

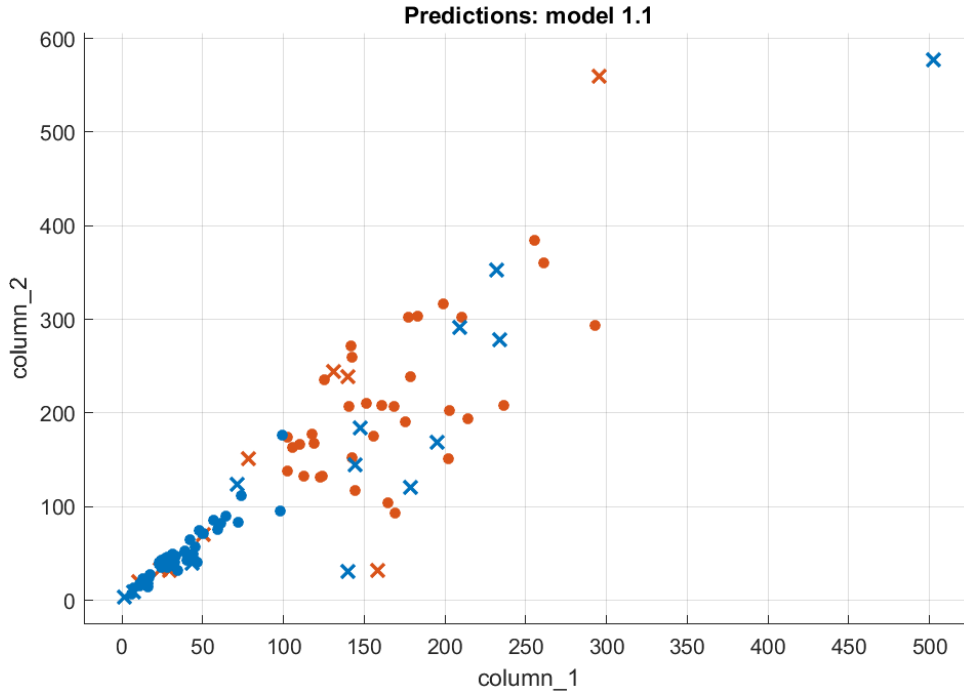
#### 4.4.2. Caso 2

En este caso se utilizan únicamente las medidas de dispersión, las de tendencia central y coeficientes de wavelet. En la figura 4.4.5 se observa la mejor precisión obtenida entre los modelos de clasificación disponibles en el aplicativo Classification Learner.



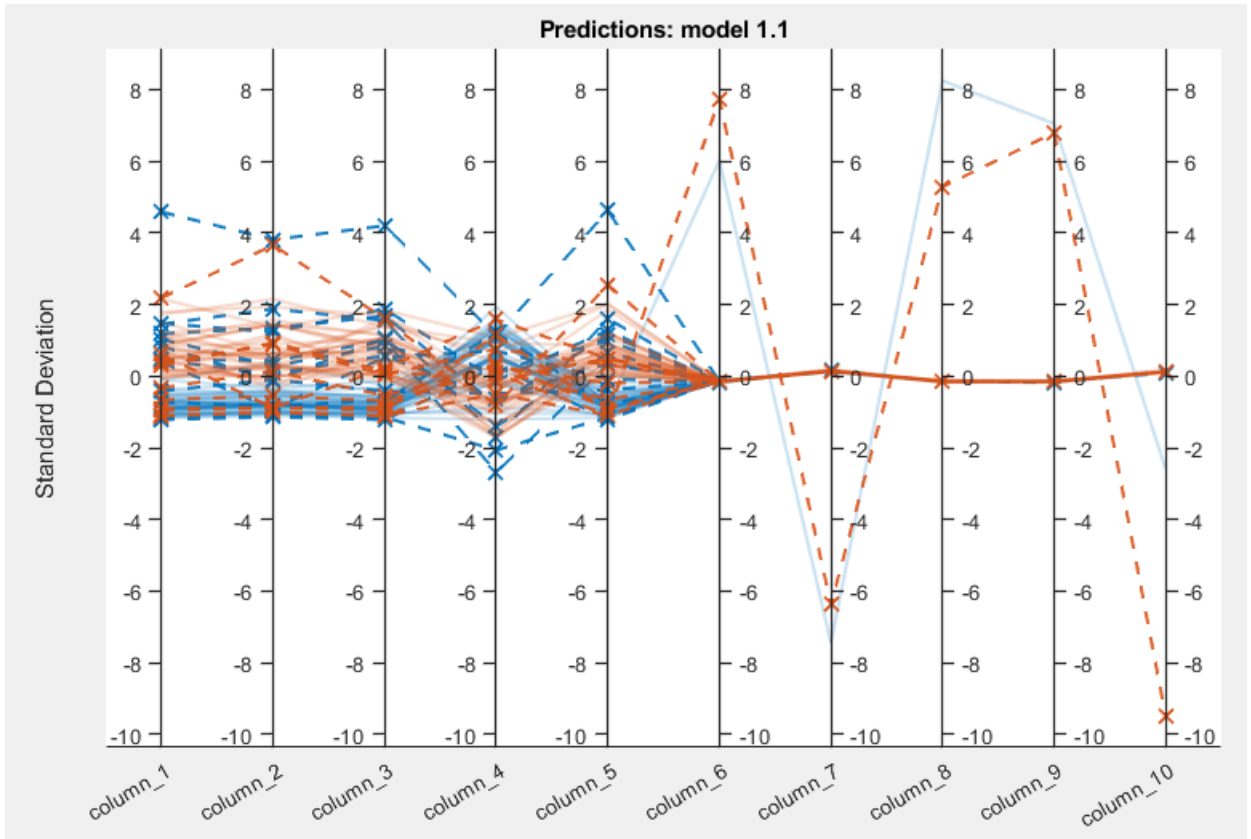
**Figura 4.4.6:** Resultado obtenido para el modelo de clasificación Linear SVM. Presenta una precisión de 92 %.

Asimismo, se presenta en los resultados la gráfica de dispersión, las coordenadas paralelas, la matriz de confusión y la región de convergencia para este caso. Obsérvese las siguientes figuras:



**Figura 4.4.7:** Gráfico de dispersión para el caso 2.

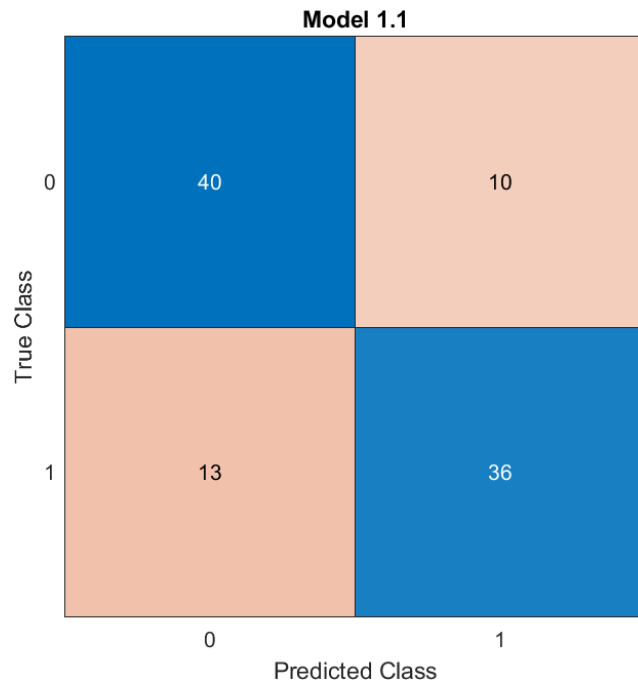
De acuerdo a la figura 4.4.7, se presentan dos características,  $column_1$  en el eje X es la desviación estándar y  $column_2$  en el eje Y representa el rango intercuartil. Los puntos de color naranja representan las arritmias auriculares y los de color azul los ritmos sinusales normales. Al igual que en el Caso 1, se destaca que los diferentes valores para las dos características presentadas están menos dispersos en las señales con ritmo sinusal normal.



**Figura 4.4.8:** Coordenadas paralelas para el caso 2.

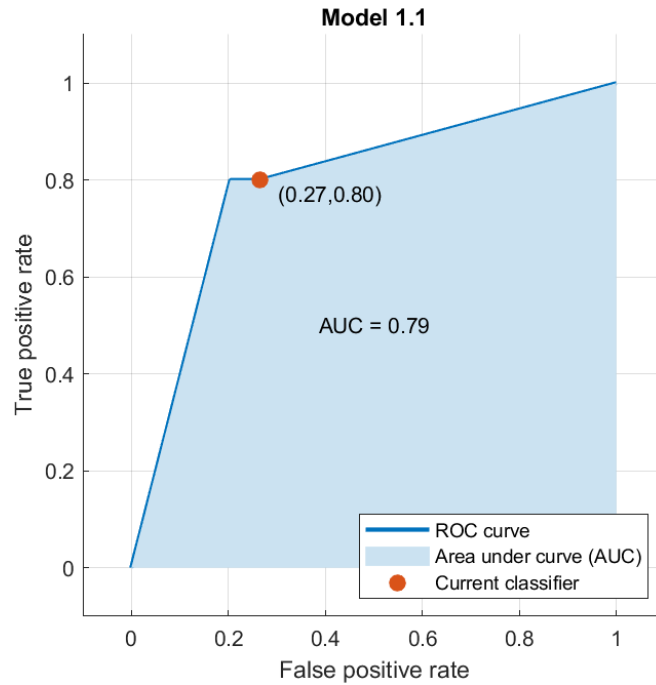
En figura 4.4.8 se presentan las coordenadas paralelas para las arritmias auriculares y los ritmos sinusales normales. Las líneas color naranja representan las arritmias auriculares, las de color azul los ritmos sinusales normales y desde  $column_1$  hasta  $column_{10}$  se muestran 10 características, las 5 primeras corresponden a las medidas estadísticas y las 5 restantes a los coeficientes de detalle wavelet. Asimismo, se observa menor relación entre las características correspondientes a las medidas estadísticas para las señales de arritmias auriculares, en comparación con las de ritmo sinusal normal. Además, las características correspondientes a los coeficientes de wavelet se observan con mayor relación para ambas clases de señales.





**Figura 4.4.9:** Matriz de confusión para el caso 2.

De acuerdo a la figura 4.4.9, se observa que 40 señales fueron correctamente clasificadas como ritmo sinusal normal y erróneamente 13 como arritmias auriculares. Asimismo, se clasificaron correctamente 36 señales como arritmias auriculares y erróneamente 10 como ritmo sinusal normal.

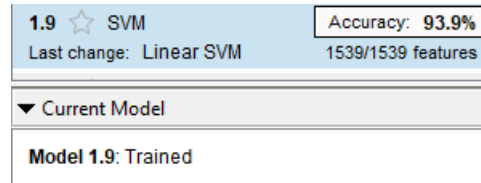


**Figura 4.4.10:** Región de convergencia para el caso 2.

La figura 4.4.10 corresponde a la región de convergencia, en la cual se observa la curva ROC en color azul oscuro, el área bajo la curva ( $AUC$ ) en color azul celeste, el punto naranja representa al clasificador, el eje X el ratio de falsos positivos ( $FPR$ ) y el eje Y el ratio de verdaderos positivos ( $TPR$ ). También, se observa que el 80 % de señales fueron correctamente clasificadas en la clase positiva frente a un 27 % que fueron clasificadas erróneamente en esta clase. También, se observa un rendimiento del 79 % para el clasificador, valor descrito por  $AUC$ .

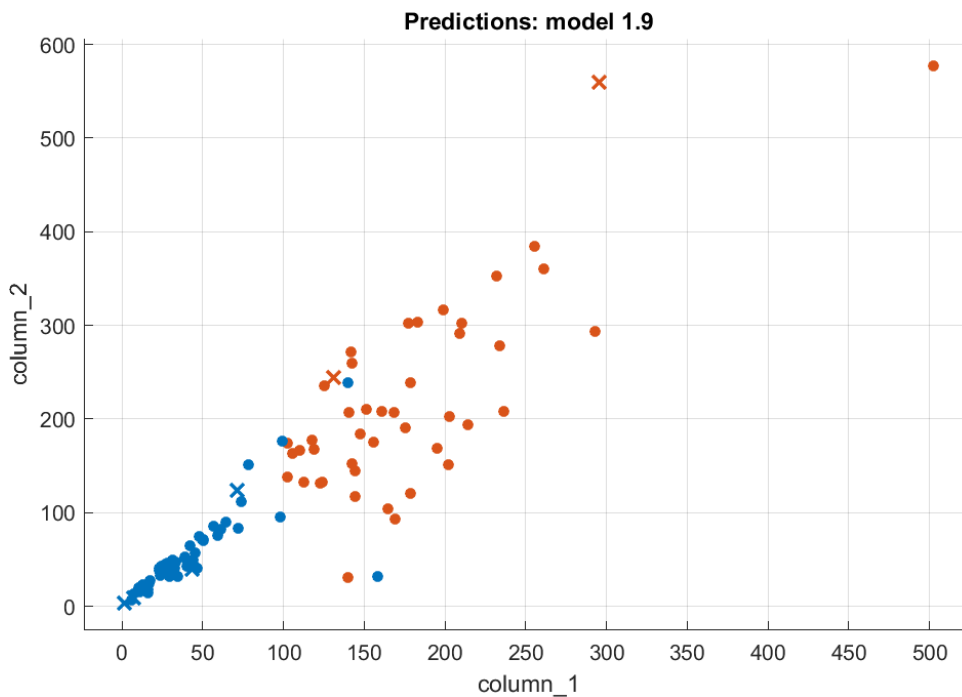
#### 4.4.3. Caso 3

Para el presente caso se utilizan las medidas estadísticas, densidad espectral, la frecuencia fundamental y coeficientes wavelet en la matriz de características. En la figura 4.4.11 se observa la mejor precisión obtenida entre los modelos de clasificación entrenados.



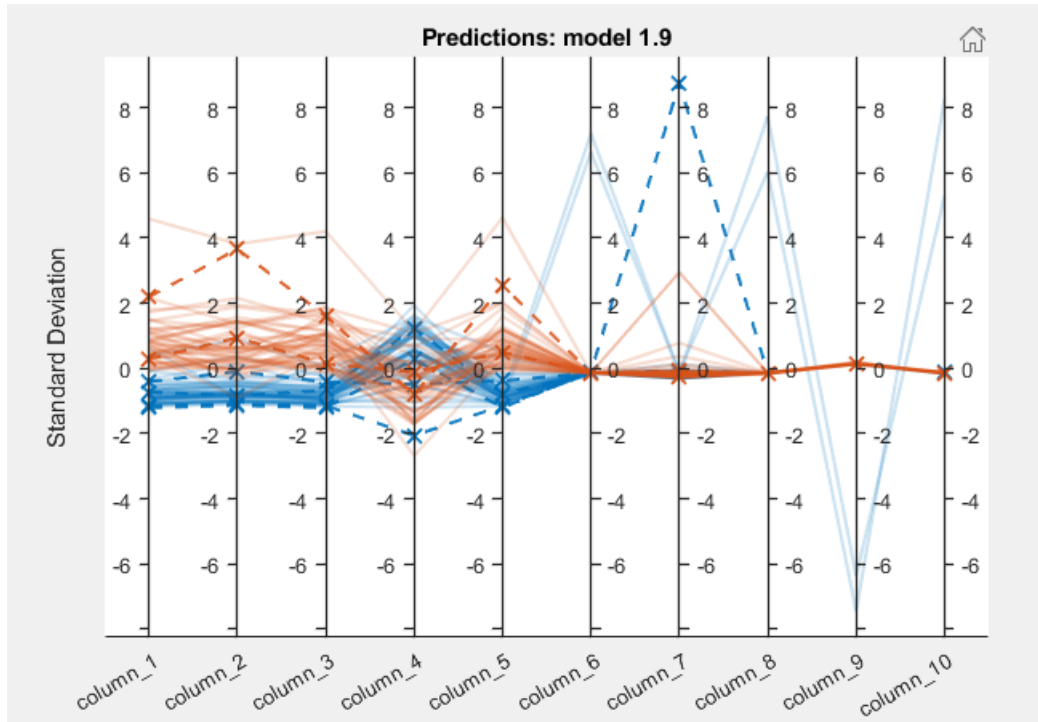
**Figura 4.4.11:** Resultado obtenido para el modelo de clasificación Linear SVM. Presenta una precisión de 94%.

Al igual que en los casos anteriores, se presentan los resultados de la gráfica de dispersión, las coordenadas paralelas, la matriz de confusión y la región de convergencia en las siguientes figuras:



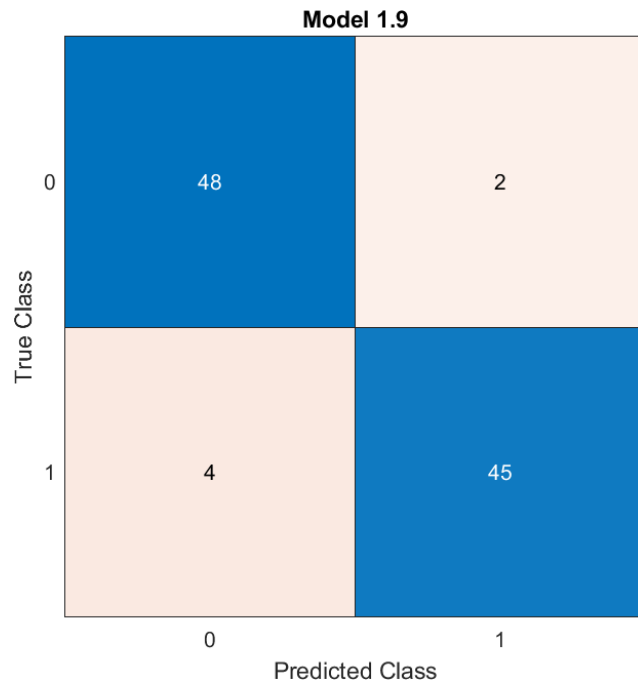
**Figura 4.4.12:** Gráfico de dispersión para el caso 3.

En la figura 4.4.12 se presentan dos características,  $column_1$  en el eje X es la desviación estándar y  $column_2$  en el eje Y representa el rango intercuartil. Los puntos de color naranja representan las arritmias auriculares y los de color azul los ritmos sinusales normales. Asimismo, se observa que las coordenadas en el plano para arritmias auriculares se encuentran más dispersas en contraste con las correspondientes a ritmo sinusal normal.



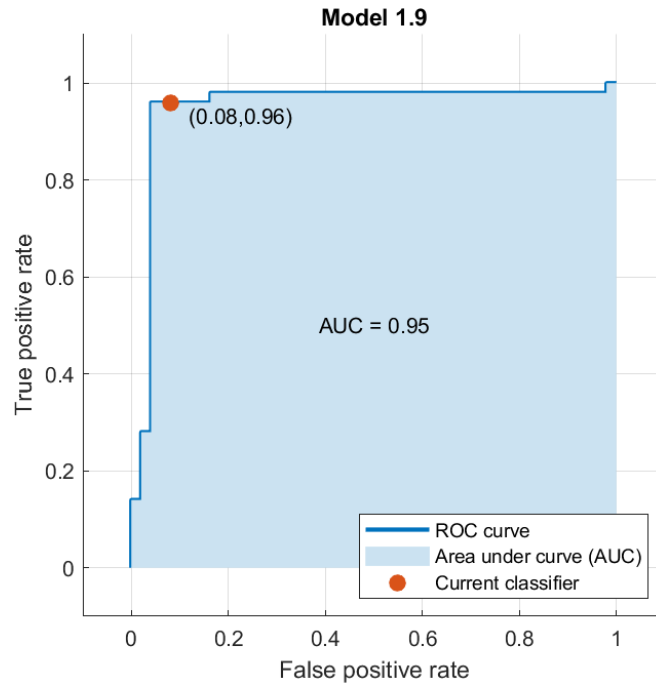
**Figura 4.4.13:** Coordenadas paralelas para el caso 3.

En la figura 4.4.13 se presentan las coordenadas paralelas para las arritmias auriculares y los ritmos sinusales normales. Las líneas color naranja representan las arritmias auriculares, las de color azul los ritmos sinusales normales. Además, para este caso desde  $column_1$  hasta  $column_{10}$  se muestran las 10 características correspondientes a las medidas de dispersión, las de tendencia central, la densidad espectral, frecuencia fundamental y coeficientes wavelet. A diferencia del caso 1 y caso 2, se observa de manera bastante notoria la poca relación existente entre los valores para las medidas estadísticas correspondientes a arritmias auriculares. Para el caso de los ritmos sinusales normales la relación entre las características es mayor.



**Figura 4.4.14:** Matriz de confusión para el caso 3.

De acuerdo a la figura 4.4.14, se observa que 48 señales fueron correctamente clasificadas como ritmo sinusal normal y erróneamente 4 como arritmias auriculares. Asimismo, se clasificaron correctamente 45 señales como arritmias auriculares y erróneamente 2 como ritmo sinusal normal.

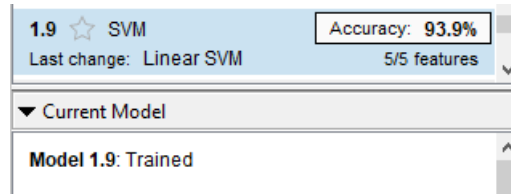


**Figura 4.4.15:** Región de convergencia para el caso 3.

La figura 4.4.15 corresponde a la región de convergencia, en la cual se observa la curva ROC en color azul oscuro, el área bajo la curva (*AUC*) en color azul celeste, el punto naranja representa al clasificador, el eje X el ratio de falsos positivos (*FPR*) y el eje Y el ratio de verdaderos positivos (*TPR*). También, se observa que el 96 % de señales fueron correctamente clasificadas en la clase positiva frente a un 8 % que fueron clasificadas erróneamente en esta clase. También, se observa un rendimiento del 95 % para el clasificador, valor descrito por *AUC*.

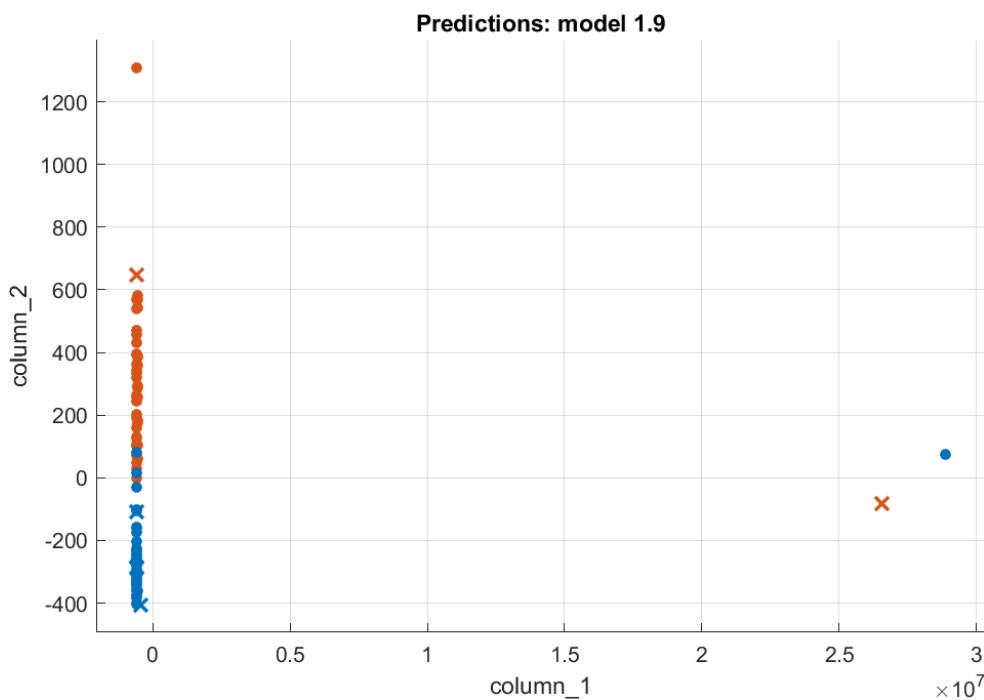
#### 4.4.4. Caso 4

Para el último caso se utilizan todas las características extraídas a través de medidas estadísticas, FFT y wavelet. Se ingresa al aplicativo Classification Learner un dataset compuesto por 5 eigenvalores obtenidos con PCA. En la figura 4.4.16 se observa el mejor valor de precisión obtenido entre los modelos de clasificación entrenados.



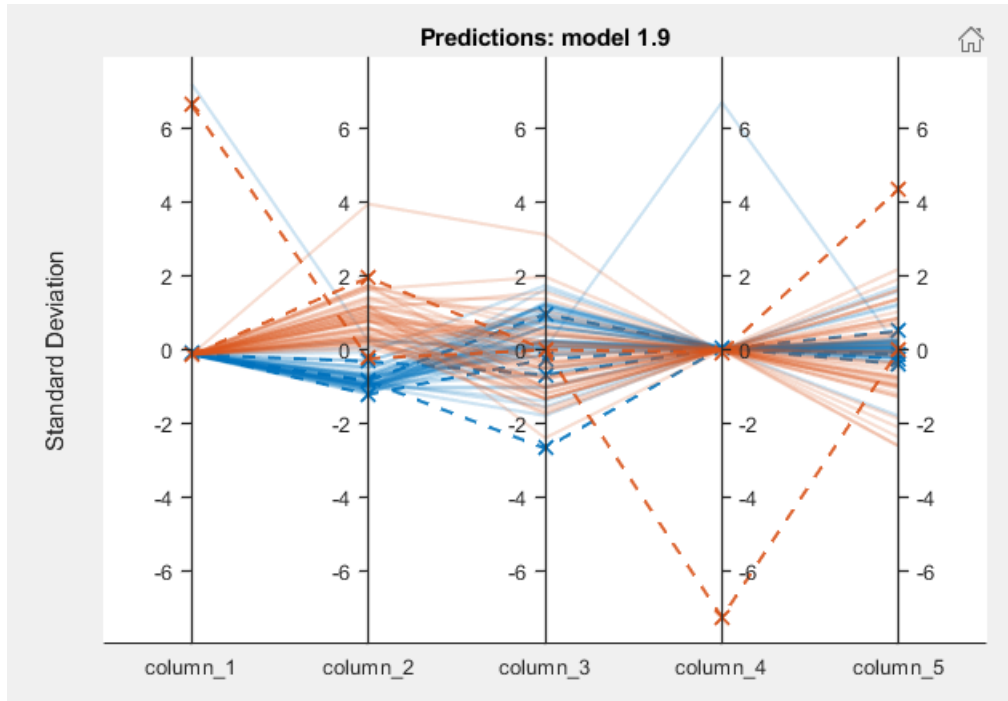
**Figura 4.4.16:** Resultado obtenido para el modelo de clasificación Linear SVM. Se obtuvo una precisión de 94 %.

Asimismo, se presentan los resultados de la gráfica de dispersión, las coordenadas paralelas, la matriz de confusión y la región de convergencia en las siguientes figuras:



**Figura 4.4.17:** Gráfico de dispersión para el caso 4.

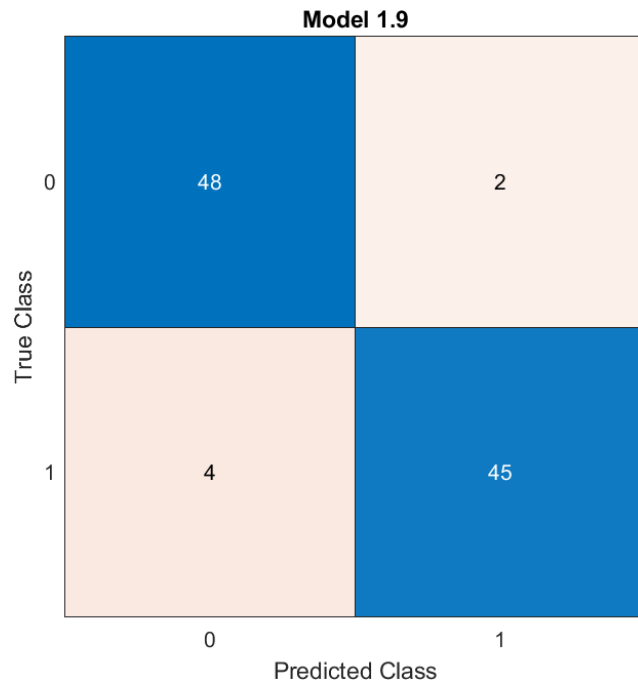
En la figura 4.4.17 se presentan dos características,  $column_1$  en el eje X representa la primera componente principal y  $column_2$  en el eje Y corresponde a la segunda componente principal obtenida con PCA. Los puntos de color naranja representan las señales de arritmias auriculares y los de color azul las de ritmo sinusal normal. Así pues, se observa al igual que en los tres casos anteriores que, existe mayor dispersión para las coordenadas de las señales de arritmias auriculares en comparación con las de ritmo sinusal normal, de acuerdo a los valores para cada componente principal.



**Figura 4.4.18:** Coordenadas paralelas para el caso 4.

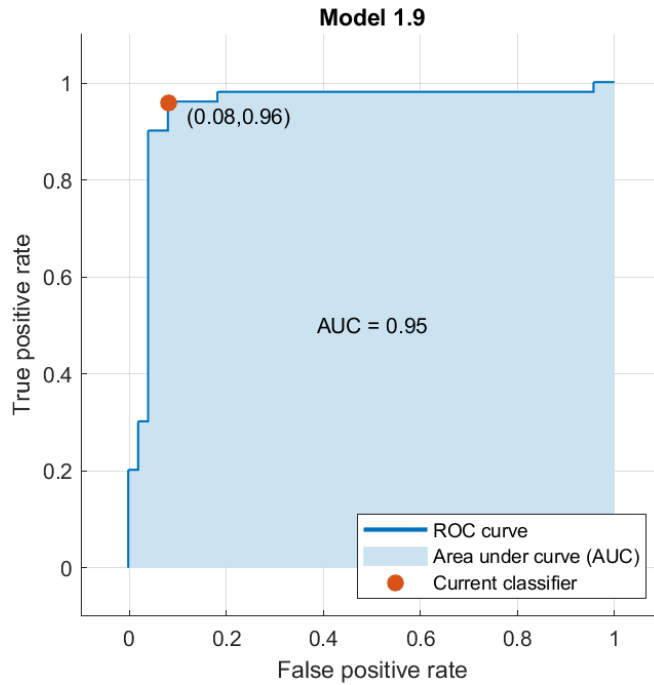
En la figura 4.4.18 se presentan las coordenadas paralelas para las arritmias auriculares y los ritmos sinusales normales. Las líneas color naranja representan las arritmias auriculares, las de color azul los ritmos sinusales normales. Además, para este caso desde  $column_1$  hasta  $column_5$  se muestran las 5 características correspondientes a los eigenvalores obtenidos con PCA. En este caso se observa menor relación entre las señales de arritmias auriculares respecto a los valores de la componente principal 2 y la componente principal 3. Lo mismo ocurre con las señales de ritmo sinusal normal. Sin embargo, la relación respecto a las componentes principales sigue siendo notoriamente menor para las arritmias auriculares.





**Figura 4.4.19:** Matriz de confusión para el caso 4.

De acuerdo a la figura 4.4.19, se observa que 48 señales fueron correctamente clasificadas como ritmo sinusal normal y erróneamente 4 como arritmias auriculares. Asimismo, se clasificaron correctamente 45 señales como arritmias auriculares y erróneamente 2 como ritmo sinusal normal.



**Figura 4.4.20:** Región de convergencia para el caso 4.

La figura 4.4.20 corresponde a la región de convergencia, en la cual se observa la curva ROC en color azul oscuro, el área bajo la curva (*AUC*) en color azul celeste, el punto naranja representa al clasificador, el eje X el ratio de falsos positivos (*FPR*) y el eje Y el ratio de verdaderos positivos (*TPR*). También, se observa que el 96% de señales fueron correctamente clasificadas en la clase positiva frente a un 8% que fueron clasificadas erróneamente en esta clase. También, se observa un rendimiento del 95% para el clasificador, valor descrito por *AUC*.

## Capítulo 5

# Análisis de resultado

Después de ejecutar los experimentos según lo formulado en la sección de metodología y obtener los resultados mostrados, la evidencia indica, además de un buen desempeño de la estructura propuesta en la metodología, desafíos y oportunidades de mejoras para trabajos futuros.

Las figuras 4.3.1, 4.3.2, 4.3.3, 4.3.4 y 4.4.1 muestran los gráfico de dispersión para ver el desempeño de las medidas de dispersión asociadas con la desviación estándar, media de la desviación absoluta, rango, rango intercuartil y la mediana respectivamente. El enfoque de las métricas de dispersión es que sean capaz de mostrar la variabilidad de la frecuencia cardíaca, ya que esta característica es diferente entre las señales de ritmo sinusal normal y las señales de arritmia cardiaca. Podemos observar que para efectos prácticos la desviación estándar, media de la desviación absoluta, rango intercuartil y rango, muestran un buen desempeño ya que el rango de dispersión se ve mayor para las señales con arritmias marcadas con estrellas rojas, mientras que las señales de ritmo sinusal normal marcadas con triángulos verdes tienden a ser más centradas a la media del conjunto de datos. Sin embargo, a excepción de lo anterior vemos que la mediana muestra un desempeño distinto y esto se debe a que esta no es una medida de dispersión si no de tendencia central, lo anterior insinúa un mejor desempeño de las medidas de dispersión que las de tendencia central al momento de cuantificar la variabilidad de la frecuencia cardíaca, lo cual era esperado debido a la naturaleza del problema.

Por otro lado, vemos que el rango de especificidad y sensibilidad cambia para cada caso en la ejecución de los experimentos lo cual nos permite cuantificar el desempeño individual de

cada conjunto de datos. La tabla 5.0.1 muestra los resultados en relación para cada caso.

Caso	SE	SP
MD-DSP-FD	91.83 %	98 %
MD-WC	73.46 %	80 %
MD-DSP-FD-WC	91.83 %	96 %
PCA	91.83 %	96 %

**Cuadro 5.0.1:** especificidad y sensibilidad de cada caso, a continuación se relacionan los acrónimos. SE=sensibilidad, SP=especificidad, MD=medidas de dispersión, DSP=densidad espectral de potencia, FD=frecuencia dominante, WC=coeficientes wavelet, PCA=análisis de componentes principales

En base a la tabla 5.0.1 podemos ver un mejor desempeño asociado al uso de las características espectrales relacionadas a la transformada de Fourier, tales como, la densidad espectral de potencia y la frecuencia dominante que usando los coeficientes de detalle wavelet. Para la combinación, medidas de dispersión, densidad espectral de potencia y frecuencia dominante correspondientes al primer caso, vemos un valor de especificidad y sensibilidad de 98% y 91,83% respectivamente. Mientras que para la combinación de medidas de dispersión junto con los coeficientes de detalle wavelet vemos valores de especificidad y sensibilidad de 80% y 73,46% respectivamente. Lo anterior representa un decremento en la especificidad y la sensibilidad de 18% y 18,37% por el uso de los coeficientes wavelet, lo que nos permite concluir un mejor desempeño de la transformada de Fourier para esta aplicación. Lo anterior se explica debido a las diferencias que existen en los espectros de las señales de ritmo sinusal normal y arritmia cardiaca, tales como, frecuencia fundamental de la componente auricular que en ritmo sinusal normal se encuentra alrededor de  $1,2Hz$  mientras que para las señales de AF y AFL comúnmente se encuentra en  $6Hz$  y  $4Hz$  respectivamente . Sumado a lo anterior, se encuentra la variación de los armónicos que en ritmo sinusal normal suelen ser de amplitud creciente y luego decreciente, mientras que en las señales de arritmia auricular se presentan con amplitud decreciente. Lo anterior confirma las relaciones encontradas en el estudio [63]. Por otro lado, el uso de los coeficientes de detalle wavelet al tercer nivel de descomposición nos permite obtener la actividad ventricular, específicamente los complejos QRS libres del resto

de componentes de ECG. Aunque existe una diferencia entre la actividad ventricular asociada a las señales de ritmo sinusal normal y la actividad ventricular de las señales con arritmia auriculares, esta diferencia está caracterizada más por la distancia de los intervalos RR en sí, ya que el impulso eléctrico que permite la contracción auricular y ventricular en el corazón, es irregular para los casos de arritmias auriculares. Basado en lo anterior, se explica que la diferencia entre señales de ritmo sinusal normal y arritmia auricular presente variaciones en los periodos de activación ventricular, más que en la forma de onda de los complejos QRS y por ende se explica el bajo desempeño de los coeficientes wavelet. Por otro lado, vemos que el uso de PCA en el último caso nos ayuda a reducir la dimensión de la matriz de características, lo cual podría ser útil para aplicaciones futuras en tiempo real ya que permite reducir el costo computacional del método, sin embargo, su desempeño es equiparable al caso 1 y 3 cuando se aplican las medidas de dispersión junto con las características espectrales de la transformada de Fourier.

Finalmente, se observa el rendimiento de los diferentes clasificadores de machine learning. Por lo general se observa un mejor rendimiento usando máquinas de soporte vectorial con un hiper-plano lineal 4.4.6, 4.4.16 o en general análisis lineales. Lo anterior se puede explicar por la distribución de los datos como se observa en las gráficas de dispersión [4.3.1 -4.4.1] donde la separación entre cada clase puede ser representada con una línea recta con pendiente positiva.

## Capítulo 6

# Conclusiones y recomendaciones

El presente trabajo muestra una metodología estructurada, efectiva para la detección automática de pacientes con arritmias auriculares. Además, se logró por medio del algoritmo de Pan-Tompkins detectar los distintos complejos QRS tanto para las señales de ritmo sinusal normal como las de arritmias auriculares. En base a lo anterior, se calcula la variabilidad de la frecuencia cardiaca como la distancia que hay entre cada pico R detectado.

Con el fin de cuantificar la variabilidad de la frecuencia cardiaca se implementaron medidas de dispersión. En base a los resultados arrojados por las medidas de dispersión, se logra inferir que los registros ECG de arritmias auriculares presentan mayor variabilidad de la frecuencia cardiaca en comparación con los ECG de ritmo sinusal normal.

Por otro lado, la transformada rápida de Fourier fue implementada para extraer características, tales como, la frecuencia dominante y la densidad espectral de potencia de cada una las señales de las bases de datos. También, se utilizó la descomposición wavelet con 3 niveles de descomposición y wavelet madre db2, para extraer los coeficientes de detalle. Los resultados arrojados nos permiten concluir para esta aplicación un mejor desempeño de las características extraídas por la transformada de Fourier que los coeficientes wavelet. Con lo anterior se concluye que para este enfoque en específico, las características extraídas con la transformada de fourier son más dicientes que los coeficientes de detalle wavelet.

El uso de medidas de dispersión y características espectrales, resulta ser efectivo en la detección automática de pacientes con arritmias auriculares apoyado por métodos de

inteligencia artificial, para ello se debe tener en cuenta el tipo de características que se extraen en base a la filosofía del problema que se aborda ya que no todas las características tendrán el mismo desempeño en todos los problemas.

Finalmente, cabe resaltar que el campo de la detección de arritmias auriculares es muy amplio y presenta un desafío actual, ya que aún hay un gran porcentaje de pacientes que no están siendo diagnosticados, para trabajos futuros se puede considerar el uso de este trabajo complementando el set de características con características relacionadas a la morfología y contenido espectral de la onda P, así como incluir etapas más robustas de pre-procesamiento para hacer el método más inmune al ruido y usando menor número de derivaciones para que pueda ser implementado en un dispositivo vestible en tiempo real.

# Bibliografía

- [1] WHO Las. principales causas de defunción [internet]. ginebra: Oms; 2019 [citado 12 ene 2019], 10.
- [2] DANE. Dane. defunciones no fetales 2019 preliminar. <https://www.dane.gov.co/index.php/estadisticas-por-tema/salud/nacimientos-y-defunciones/defunciones-no-fetales/defunciones-no-fetales-2019>, 31(4), 2011.
- [3] Christopher RC Wyndham. Atrial fibrillation: the most common arrhythmia. *Texas Heart Institute Journal*, 27(3):257, 2000.
- [4] Victor Manuel Castillo Araujo. Big data y el camino a la inteligencia artificial autónoma big data and the road towards and autonomic artificial intelligence.
- [5] Edgardo Escobar. Desafíos de la cardiología: desde la investigación básico-clínica a la era digital. *Revista chilena de cardiología*, 37(1):58–67, 2018.
- [6] Marian y Piela Michal y Castells Francisco y Leksinski JM y Contreras-Ortiz Sonia H Giraldo-Guzmán, Jader y Kotas. Aplicación de filtrado espacio-temporal para mejorar las formas de onda de la actividad auricular. In *Actas de la 2da Conferencia Internacional sobre Sensores, Procesamiento de Señales e Imágenes de 2019*.
- [7] Lasse Rouhiainen. Inteligencia artificial. *Madrid: Alienta Editorial*, 2018.
- [8] Jiapu Pan and Willis J Tompkins. A real-time qrs detection algorithm. *IEEE transactions on biomedical engineering*, (3):230–236, 1985.
- [9] Raaed Faleh Hassan and Sally Abdulmunem Shaker. Ecg signal de-noising and feature extraction using discrete wavelet transform. *International Journal of Engineering Trends and Technology*, 63(1):32–39, 2018.
- [10] C Vidal and L Pavesi. Desarrollo de un sistema de adquisición y tratamiento de señales electrocardiográficas, universidad de tarapacá.
- [11] Ana Vivian Calderón Echeverría, Mateo Navas Luquez, Carolina Cristancho Lenis, and Juan Manuel López López. Extracción del ecg fetal en electrocardiografía no invasiva por medio de señal de referencia. *Encuentro Internacional de Educación en Ingeniería*, 2019.
- [12] Nissrin Amrani El Yaakoubi et al. Procesamiento del complejo qrs característico del electrocardiograma (ecg). 2020.
- [13] D Cuesta Frau, D Novák, JC Pérez Cortés, G Andréu García, V Eck, C Sastre Mengual,



- and M Llorca Alcón. Reducción del ruido en señales electrocardiográficas mediante la transformada wavelet. *Energy*, 200:250, 2000.
- [14] Victoria Eugenia Montes, Gustavo A Guarín, and Germán Castellanos Domínguez. Extracting ecg signal characteristics based on non-linear transformations and wavelets. *Ingeniería e Investigación*, 25(3):39–48, 2005.
- [15] Harold Santiago Blanco Ortega. Clasificación de arritmias cardiacas empleando la transformada wavelet y redes neuronales. 2006.
- [16] Wilmer H Rojas, Jhon F Herrera, Jadir G Acuña, Pablo A Muñoz, and Francisco J Ibarгүйen. Análisis y supervisión de la señal cardiaca con transmisión inalámbrica de datos. *Scientia et technica*, 1(1):151–156, 2007.
- [17] Angela Patricia Cabarcas Barboza and Tania Yineth Guerrero Castilla. Técnicas de extracción de características de señales biomédicas. 2007.
- [18] Emran M Abu Anas, Soo Y Lee, and Md K Hasan. Sequential algorithm for life threatening cardiac pathologies detection based on mean signal strength and emd functions. *Biomedical engineering online*, 9(1):1–22, 2010.
- [19] María Cristina Pinzón Duque. *Análisis de señal del impulso cardíaco para el mejoramiento del diagnóstico de patologías del corazón*. PhD thesis, Universidad Tecnológica de Pereira. Facultad de Ciencias Básicas. Maestría . . . , 2011.
- [20] Aya F Khalaf, Mohamed I Owis, and Inas A Yassine. A novel technique for cardiac arrhythmia classification using spectral correlation and support vector machines. *Expert Systems with Applications*, 42(21):8361–8368, 2015.
- [21] Fernando Riveros Sanabria et al. *Análisis con Wavelets de alteraciones electrocardiográficas en pacientes Chagásticos crónicos*. PhD thesis, Universidad EAFIT, 2012.
- [22] Muhidin Mohamed and Mohamed Deriche. An approach for ecg feature extraction using daubechies 4 (db4) wavelet. *International Journal of Computer Applications*, 96(12):36–41, 2014.
- [23] Diego Alejandro Tovar Salazar, Alejandro José Orozco Naranjo, Pablo Andrés Muñoz Gutiérrez, Héctor Murillo Wills, and Javier Alejandro Granada. Transformada wavelet y máquinas de soporte vectorial para la identificación de arritmias wavelet transform and support vector machines for the arrhythmia identification.
- [24] Javier E González Barajas. Cálculo del umbral para detección de la onda r del complejo cardiaco. *TecnoLógicas*, 17(32):47–55, 2014.
- [25] Gopisetty Ramesh, DD Satyanarayana, and M Sailaja. Automatic detection of cardiac arrhythmia through ecg signal analysis: A review\*. *vol*, 7:1–12, 2017.
- [26] OSCAR EDUARDO VERA, EDISON DUQUE CARDONA, and JORGE RIVERA PIEDRAHITA. Extracción de características de la señal electrocardiográfica mediante software de análisis matemático. *Scientia et technica*, 12(31):59–64, 2006.
- [27] Jesús Iván Lara Prado. El electrocardiograma: una oportunidad de aprendizaje. *Revista de la Facultad de Medicina (México)*, 59(6):39–42, 2016.

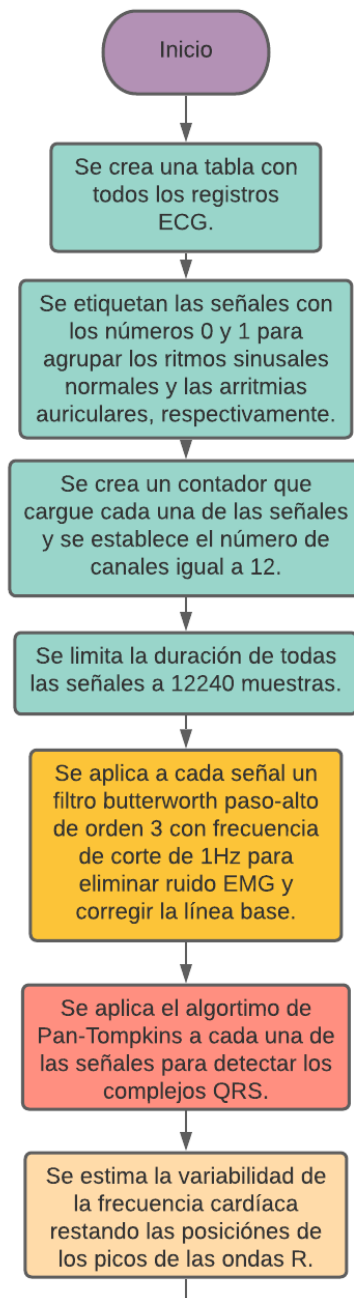
- [28] Luis Felipe Paz Campos and Luis Isaac Herrera Velásquez. *Diseño y construcción de un sistema que simule las señales eléctricas del corazón*. PhD thesis, Universidad Nacional de Ingeniería, 2017.
- [29] Alberto Ochoa, Marcelo Maciel, Felipe Estrada, Columba Díaz, Ramón Félix, José Alvarez, José Clemente VÁSQUEZ, and Centro Universitario de Investigaciones Biomédicas. Sistema de adquisición y procesamiento de señales electrocardiográficas. *Sistemas Cibernética e informática*, 7:1–15, 2010.
- [30] César Enrique Rose-Gómez and María Trinidad Serna-Encinas. Procesamiento del electrocardiograma para la detección de cardiopatías. *Researchgate. Net*, no. May, pages 3–6, 2015.
- [31] Aswathy Velayudhan and Soniya Peter. Noise analysis and different denoising techniques of ecg signal-a survey. *IOSR journal of electronics and communication engineering*, 3:641–644, 2016.
- [32] Willis J Tompkins. Biomedical digital signal processing. *Editorial Prentice Hall*, 1993.
- [33] Kenneth M y Miller Martin H Holsinger, William P y Kempner. Un preprocesador qrs basado en la diferenciación digital. *IEEE Transactions on Biomedical Engineering*.
- [34] Investigación cuantitativa de las reglas de detección de qrs utilizando la base de datos de arritmias mit / bih. *transacciones IEEE sobre ingeniería biomédica*.
- [35] Philip De Chazal, Maria O’Dwyer, and Richard B Reilly. Automatic classification of heartbeats using ecg morphology and heartbeat interval features. *IEEE transactions on biomedical engineering*, 51(7):1196–1206, 2004.
- [36] Philip De Chazal and Richard B Reilly. A patient-adapting heartbeat classifier using ecg morphology and heartbeat interval features. *IEEE transactions on biomedical engineering*, 53(12):2535–2543, 2006.
- [37] Álvaro Montenegro et al. Análisis espectral. Technical report, Universidad Javeriana-Bogotá, 2009.
- [38] I ROJAS SOLORZANO. *Herramienta de Análisis Espectral*. PhD thesis, 2013.
- [39] Harry Dym et al. Fourier series and integrals. 1972.
- [40] P Gomez, J Bobadilla, and J Bernal. La transformada de fourier, una vision pedagogica.
- [41] William T Cochran, James W Cooley, David L Favon, Howard D Helms, Reginald A Kaenel, William W Lang, George C Maling, David E Nelson, Charles M Rader, and Peter D Welch. What is the fast fourier transform? *Proceedings of the IEEE*, 55(10):1664–1674, 1967.
- [42] C Galdo and P Cetta. La transformada discreta de fourier. *Apuntes del Seminario de Matemática Aplicada a la Composición Musical. FACM-UCA*, 2009.
- [43] Diana Zoraida Briceño Rodríguez et al. Herramienta para la medición de ocupación de canales para radio cognitiva en el rango de 470 mhz a 900 mhz.
- [44] Ana Lucía Schmidt. Fft: Transformada rápida de fourier. *Estudiante de Ingeniería en Sistemas de Computación. Universidad Nacional de Sur. Bahía Blanca, Argentina. Marzo 2013*, 2013.

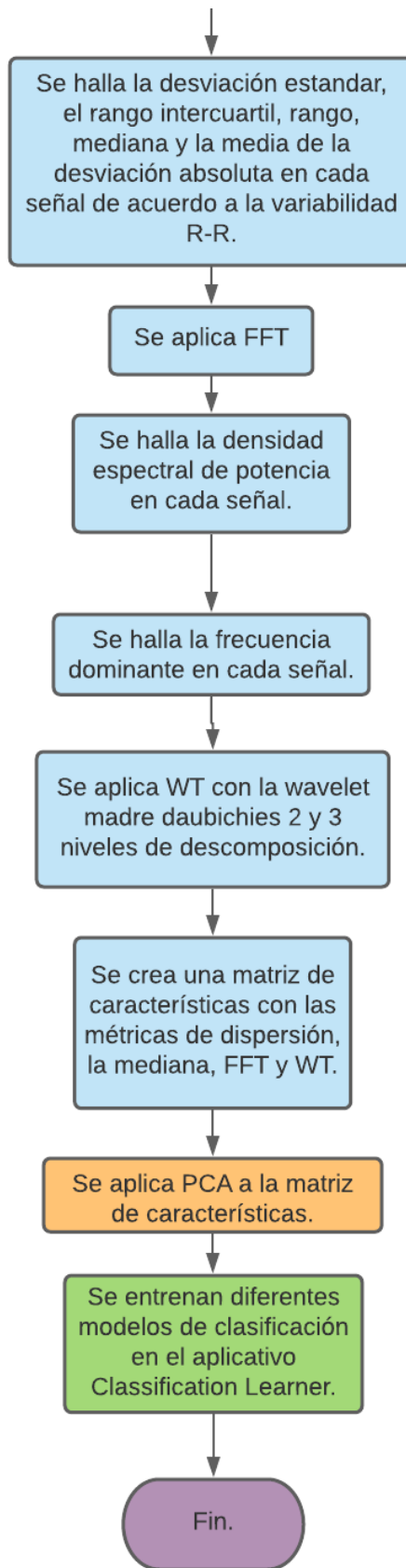
- [45] Natalia Nieto and Diana Marcela Orozco. El uso de la transformada wavelet discreta en la reconstrucción de señales senosoidales. *Scientia et technica*, 14(38):381–386, 2008.
- [46] AA Chanerley, NA Alexander, and B Halldorsson. On fling and baseline correction using quadrature mirror filters. In *12th International Conference on Civil, Structural and Environmental Engineering Computing*, 2009.
- [47] S Guarín et al. Estadística aplicada, 2002.
- [48] Thomas Hardy. Ia (inteligencia artificial). *Polis: Revista Latinoamericana*, (2):18, 2001.
- [49] Neilys González Benítez, Vivian Estrada Sentí, and Aylin Febles Estrada. Estudio y selección de las técnicas de inteligencia artificial para el diagnóstico de enfermedades. *Revista de Ciencias Médicas de Pinar del Río*, 22(3):534–544, 2018.
- [50] ALFONSO GALIPIENSO, MARIA ISABEL, Miguel Angel Cazorla Quevedo, Otto Colomina Pardo, FRANCISCO ESCOLANO RUIZ, and MIGUEL ANGEL LOZANO ORTEGA. *Inteligencia artificial: modelos, técnicas y áreas de aplicación*. Editorial Paraninfo, 2003.
- [51] M Sc Luis Javier Mena Camaré. *Aprendizaje automático a partir de conjuntos de datos no balanceados y su aplicación en el diagnóstico y pronóstico médico*. PhD thesis, Instituto Nacional de Astrofísica, Óptica y Electrónica, 2008.
- [52] Fabio A González. Machine learning models in rheumatology. *Revista Colombiana de Reumatología*, 22(2):77–78, 2015.
- [53] Yoshua Bengio, Ian Goodfellow, and Aaron Courville. *Deep learning*, volume 1. MIT press Massachusetts, USA:, 2017.
- [54] Dawn E Holmes and Dulcinea Otero-Piñeiro. *Big Data: una breve introducción*. Antoni Bosch editor Barcelona, 2018.
- [55] Viktor Mayer-Schönberger and Kenneth Cukier. *Big data: la revolución de los datos masivos*. Turner, 2013.
- [56] Yuri Demchenko, Cees De Laat, and Peter Membrey. Defining architecture components of the big data ecosystem. In *2014 International Conference on Collaboration Technologies and Systems (CTS)*, pages 104–112. IEEE, 2014.
- [57] Anthony Chang. Big data en medicina: La inteligencia artificial que viene.
- [58] Bani Bansal, Malti y Gandhi. Iot big data en smart healthcare (monitorización de ecg). In *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*.
- [59] R Varatharajan, Gunasekaran Manogaran, and MK Priyan. A big data classification approach using lda with an enhanced svm method for ecg signals in cloud computing. *Multimedia Tools and Applications*, 77(8):10195–10215, 2018.
- [60] F Costa, M Fineschi, P Giudice, G Lavigna, C Selmo, and M Risk. Evaluación del preprocesamiento con acp y tdf en el análisis de ecg utilizando una red neuronal.
- [61] J Amat. Análisis de componentes principales (principal component analysis, pca) y t-sne, 2017.

- [62] Santiago De la Fuente. Análisis componentes principales. *Recuperado en*, 20(2):2015, 2011.
- [63] Rieta' JJ, J Millet-Roig, V Zarzoso, F Castells, Sánchez, R García-Civera, and S Morell. Atrial fibrillation, atrial flutter and normal sinus rhythm discrimination by means of blind source separation and spectral parameters extraction.



# Anexos





En las imágenes anteriores se observa el diagrama correspondiente al software desarrollado.

Características														
99x1539 double														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	261.4596	360	855	735.5000	208.7755	8.2847e+03	17.8922	0.0013	-0.0075	-0.0095	0.0213	0.0050	-0.0088	-0.
2	140.8426	207	484	821	113.3571	8.5228e+03	1.3889	8.2421e-04	0.0011	-0.0105	-0.0077	0.0106	-0.0011	-0.
3	131.0834	244	401	767.5000	116.7857	1.3293e+04	4.6569	-0.0019	-0.0065	-3.8760e-04	-0.0013	-0.0029	-0.0047	-0.
4	45.0314	57	144	821.5000	34.0408	2.4784e+04	3.8399	0.0010	-0.0030	-0.0170	9.7674e-04	0.0100	-0.0077	0.
5	11.1871	17	39	982	8.8333	1.6431e+04	4.1667	-0.0011	-0.0017	-0.0026	-0.0012	3.4818e-04	-0.0021	-0.
6	48.1768	74.5000	159	988	36.9091	1.3631e+04	4.0850	5.8555e-04	-2.2233e-04	-0.0020	-0.0049	0.0033	0.0049	-0.
7	33.4812	47.5000	97	965	27.0972	1.3605e+04	1.1438	-0.0018	-0.0079	-0.0047	0.0021	0.0011	-0.0095	0.
8	32.7729	41	109	994.5000	25.3889	1.3377e+04	3.1046	2.2994e-04	-0.0090	0.0064	0.0122	-0.0021	0.0108	-0.
9	49.7577	71.5000	169	875	41.5385	2.3422e+04	1.2255	-8.2317e-04	0.0034	0.0085	0.0127	0.0038	-0.0150	-0.
10	40.2306	43	118	1.1385e+03	29.2000	1.1649e+04	1.8791	3.4237e-04	-0.0020	3.8452e-04	-0.0049	0.0061	-0.0031	0.
11	99.0085	176.2500	286	971	84.1157	1.3506e+04	2.2876	3.2640e-04	0.0030	0.0035	-3.5961e-04	-0.0037	0.0045	-0.
12	73.6297	111.5000	252	961.5000	61.2222	1.4235e+04	2.1242	9.7001e-04	0.0029	0.0010	0.0069	0.0013	0.0094	5.2014
13	98.0117	95	321	1002	71.3719	1.2248e+04	1.9608	-4.0626e-04	0.0103	-0.0069	-0.0027	-0.0072	0.0083	0.
14	59.3411	76	210	1022	46.5289	1.2238e+04	2.9412	1.6724e-04	-0.0035	-0.0042	9.9915e-05	0.0033	0.0042	-0.
15	41.8501	65	136	1.0565e+03	33.7200	1.2479e+04	1.9608	0.0021	7.9964e-04	-0.0042	8.9322e-04	0.0024	0.0025	-1.5250
16	24.0650	35	78	1.0865e+03	19.9000	1.1060e+04	3.7582	-7.1502e-05	-0.0104	0.0098	-0.0015	0.0066	0.0170	-0.
17	61.3870	82.5000	186	1046	49.2231	1.0739e+04	2.9412	-1.6559e-04	-1.9901e-04	0.0084	-0.0014	-2.6350e-04	-0.0042	-0.
18	16.4201	14.7500	64	601	11.7618	1.6681e+04	1.7157	2.5430e-04	-0.0023	-0.0020	0.0032	0.0061	-0.0030	-0.
19	295.4644	559.7500	828	639	255.8947	1.0005e+04	1.9608	2.1107e-04	0.0046	0.0014	-0.0170	-0.0043	0.0047	0.
20	78.3159	151	202	837	69.7143	1.2440e+04	3.3497	0.0011	-0.0017	0.0092	6.8131e-04	-0.0012	0.0053	-0.

En la imagen anterior se muestra la matriz de características para el caso 3. Las columnas equivalen a las diferentes características, solo se lograron capturar graficamente 14, teniendo en cuenta que es una matriz de  $99 \times 1539$ . Las primeras 7 columnas corresponden respectivamente a: desviación estandar, rango intercuartil, rango, mediana, media de la desviación absoluta, densidad espectral de potencia y frecuencia dominante. Las columnas restantes son los coeficientes wavelet. Asimismo, las filas corresponden al número de señales ECG, de las cuales 50 pertenecen a ritmo sinusal normal y 49 son de arritmias auriculares.

Para los casos 1 y 2, la matriz de características varió únicamente en el número de columnas, es decir el número de características.



	1	2	3	4	5	6	7
1	-5.8343e+05	567.6872	81.3499	-8.5415	48.1733		
2	-5.8320e+05	130.5463	64.6105	-1.5159	26.3300		
3	-5.7843e+05	80.8557	-3.2809	-1.2454	90.2277		
4	-5.6694e+05	-249.4555	-25.2389	4.7801	-1.6330		
5	-5.7529e+05	-403.6088	102.7338	6.8955	-2.2892		
6	-5.7809e+05	-268.6800	140.9545	4.5574	12.0333		
7	-5.7811e+05	-330.9034	102.6519	5.6850	7.9449		
8	-5.7834e+05	-329.9152	133.3039	5.7072	-2.3729		
9	-5.6830e+05	-233.1879	33.4840	4.3500	4.6329		
10	-5.8007e+05	-352.4837	276.0533	5.6861	-0.3470		
11	-5.7821e+05	-101.4434	162.7720	1.3856	67.7879		
12	-5.7748e+05	-159.4531	139.8098	2.8700	16.0327		
13	-5.7947e+05	-106.1636	195.2025	2.2849	-20.8293		
14	-5.7948e+05	-227.3333	186.0908	4.0051	-2.8292		
15	-5.7924e+05	-309.7104	201.5890	5.0721	12.3053		
16	-5.8066e+05	-383.3340	214.9549	6.2823	3.8667		
17	-5.8098e+05	-250.6706	205.1074	4.1451	13.1616		
18	-5.7504e+05	-291.9313	-261.8710	6.2897	-19.2217		
19	-5.8171e+05	646.1436	-2.2761	-11.6217	248.6239		
20	-5.7928e+05	-158.7042	11.3533	2.6158	70.4567		

En la imagen anterior, se observa el dataset creado a partir de PCA, correspondiente al caso 4. Al igual que en el resto de casos, las columnas corresponden al número de características y las filas al número de señales ECG. En este caso se observa una matriz de  $99 \times 5$ , ya que las 5 características de este caso son los 5 eigenvalores de las 99 señales.