



INSTITUTO POLITÉCNICO
NACIONAL



UNIDAD PROFESIONAL INTERDISCIPLINARIA EN INGENIERÍA Y
TECNOLOGÍAS AVANZADAS

TRABAJO TERMINAL II

Caracterización de precursores para patologías cardíacas mediante análisis complejo.

Autores:

Cruz Reyes Juan Manuel
Portilla Brambila Sebastian
Ramírez Palomino Alan Jesús

Asesores:

Dr. Guzmán Vargas Lev
M en C. Anzueto Rios Alvaro

Ing. Biónica

25 de enero de 2021

Agradecimientos

Al Instituto Politécnico Nacional, nuestra casa de estudios que desde el inicio nos invitó mediante su lema “LA TÉCNICA AL SERVICIO DE LA PATRIA” y sus ideales a dar lo mejor de nosotros por una causa mayor a nuestro beneficio personal. La remembranza de las horas invertidas en sus espacios y actividades formativas nos commueven al más profundo agradecimiento con nuestra Alma Mater, de la cual recibimos siempre tanto a cambio de tan poco.

A la Unidad Profesional Interdisciplinaria en Ingeniería y Tecnologías Avanzadas (nuestra UPIITA). Siempre amada y temida en iguales proporciones, por enseñarnos a su manera a ganarnos el conocimiento a pulso, y a atesorar los pequeños momentos que eluden, la gran mayoría de las veces, la mente del estudiante preocupado. Nos llevamos en el corazón entrañables recuerdos de amigos, profesores y lugares. Jamás creímos posible extrañar tanto desayunar en la cafetería.

A nuestros asesores, el Dr. Lev Guzmán Vargas, y el M. en C. Alvaro Anzueto Rios por abrirnos las puertas a un mundo nuevo y desconocido desde la instrucción que nos dieron como alumnos de la UPIITA. Agradecemos su valiosa ayuda, el aporte que sus perspectivas y conocimientos nos brindaron además de sus consejos y su tiempo sin los cuales este proyecto no sería lo mismo.

A nuestros colaboradores del Laboratorio de Sistemas Complejos, quienes siempre nos han brindado su apoyo y nos han hecho sentir bienvenidos.

Cruz Reyes Juan Manuel

A mi madre, a quien me enseño el amor por la vida y con un ejemplo de lucha, a nunca darme por vencido frente a circunstancias adversas, por todo su cariño y sus esfuerzos realizados a lo largo de su vida para que pudiese llegar a donde lo hice, por sus continuos consejos y enseñanzas que perdurarán por siempre en mi persona.

A mi padre, por siempre acompañarme en cada momento importante, por ser un gran ejemplo para mí, por inculcarme el valor de la educación y el aprendizaje, por todos los sacrificios que ha hecho constantemente para brindarme todos los recursos a su alcance y hoy poder concluir mi carrera.

A mi tía Victoria, por contar siempre con su apoyo y sus palabras de aliento, por todos sus consejos, cuidados y cariño brindado a través de todo este tiempo.

A mis abuelitos Heriberto, Margarita y Sofía, por su gran afecto, por transmitirme grandes lecciones de vida y por demostrarme el gran valor que tiene la familia.

A todos mis primos y tíos, a quienes les agradezco tantas experiencias y recuerdos vividos en familia, por todo su apoyo brindado y consejos otorgados a lo largo de mi vida.

A todos mis profesores, por transmitirme todos estos años conocimientos valiosos y por todo el tiempo dedicado en responder dudas y mostarme mis aciertos y errores, por tantas horas de clase invertidas en

mi formación, agradezco a todos.

A mis amigos, que fueron un pilar importante en esta etapa, con los que conviví y disfrute una gran parte de mi tiempo en la escuela, con pláticas extensas e interesantes, por todos los proyectos realizados en su compañía y por ser una gran fuente de motivación y aliento en momentos de frustración.

A todos ellos gracias por todo lo que me han dejado.

Portilla Brambila Sebastian

A mi padre y madre, quienes me han apoyado con sus esperanzas y visión puestas en mí sin importar lo que sucediera, poniendo su total confianza en que con mi desarrollo académico y profesional llegaría al éxito. Quienes a su forma y con sus experiencias me han enseñado lo que es lanzarte por tu vocación con toda la actitud y mantener viva esa llama no importa cuantos obstáculos se presenten. Quienes también han invertido de muchas maneras en convertirme en la persona que soy ahora.

A mis amorosas abuelas, quienes simplemente eran incapaces de concebir otra cosa que no fueran logros en lo que me propusiera, quienes invirtieron todo su cariño en mi hermana y en mí siempre, y quienes ya estaban orgullosas de mí por toda mi trayectoria antes siquiera de saber por dónde iría. Aunque ya no esté mi abuela Lulú para presenciar mi graduación, siento plenamente su afecto aún tras el tiempo recorrido, tan cercano y presente como el de mi abuela Betty.

A mis compañeros de trabajo claramente, sin quienes no sólo no hubiera podido surgir este proyecto y realizarlo durante una pandemia, sino que también han sido un gran apoyo en mi crecimiento personal y con quienes comparto un hambre por el futuro al cual podremos conquistar.

R. Palomino Alan Jesús

A mi madre, Nury. El ejemplo claro de que la vida puede hacer con las experiencias más amargas a las personas más dulces. Tú sacrificio por nosotros y tu incansable dedicación son el motor que motiva mi vida.

A mi hermana, Eri, la hija que la vida me dio sin necesidad de tener una propia. Cada risa y cada lágrima que hemos compartido es un valioso tesoro en mi corazón. Sigue siempre luchando por aquello que anhelas y verás que vas a conseguirlo. Cuentas conmigo a cada paso por siempre.

A mi incondicional Gaby Montero, aún después de 13 años de conocernos sigues sorprendiéndome cada día. Siempre me has impulsado a ser un mejor ser humano y ahora a realizar un sueño en mi vida profesional. Puedo afirmar, sin temor a equivocarme que eres y seguirás siendo el amor de mi vida. Gracias por creer en mí aún esos días en los que ni yo mismo puedo hacerlo. Te amo más.

A mi abuela, la profesora Laura López Anguiano, hoy no estás aquí para verlo, siguiendo tu consejo he llegado hasta aquí y pero puedo asegurarte que lo único verdaderamente difícil fue hacerlo sin ti. La vida me negó la oportunidad de cuidarte por más tiempo, pero siempre te llevo en mi corazón cuidando cada cariño y cada consejo como si de ti se tratara.

A mis compañeros y amigos, ustedes han sido una parte fundamental de este viaje, y agradezco infinitamente cada muestra de amistad sincera que de ustedes he recibido. Espero que el camino continúe y sepan que en mí siempre habrá una mano para ayudar.

Glosario

- algoritmo** Conjunto ordenado de operaciones sistemáticas que permite hacer un cálculo y hallar la solución de un tipo de problemas.. 16
- ambigüedad** Comportamiento, hecho, palabra o expresión que puede entenderse o interpretarse de diversas maneras.. 23
- asimetría** Desigualdad entre las partes de un todo.. 18
- bimodal** Asociación de dos modalidades: signada y hablada. 19
- biomarcador** Indicador del estado de salud, de la esperanza de vida o del riesgo de enfermedad. 16
- cardiovascular** Del corazón y los vasos sanguíneos o relacionado con ellos.. 16
- caótica** Que es absolutamente desordenado o confuso.. 18
- curtosis** Medida estadística que determina el grado de concentración que presentan los valores de una variable alrededor de la zona central de la distribución de frecuencias.. 17
- difusión** Propagación de moléculas, solutos, solventes, noticias, informaciones, pensamientos, tendencias o culturas de un medio a otro.. 33
- electrocardiograma** Procedimiento simple e indoloro que mide la actividad eléctrica del corazón.. 16
- electrodo negativo** Denominado electrodo indiferente. 33
- entropía** Medida de variabilidad o desorden de un sistema. 8, 40
- espectral** Perteneciente o relativo al espectro. 18
- fibrilación auricular** frecuencia cardíaca irregular y a menudo rápida que ocurre cuando las dos cavidades superiores del corazón presentan señales eléctricas caóticas. 18
- fluctuación** Variación de intensidad, de medida o de calidad.. 16, 19
- fractal** Objeto geométrico en el que se repite el mismo patrón a diferentes escalas y con diferente orientación. 16
- haz auriculoventricular** Formación intracardíaca consistente en un fino cordón de naturaleza muscular,. 30
- intervalo** Porción de tiempo o de espacio cuya extensión se expresa y en la cual sucede o se da una cosa.. 18
- isquemia** Reducción del flujo sanguíneo al corazón que impide al músculo cardíaco recibir suficiente oxígeno.. 16
- lineal** Toda aquella relación que evidencia una correspondencia uno a uno; Referente a una línea.. 17

- Machine Learning** Paradigma computacional que incluye un conjunto de métodos que permiten a las computadoras “aprender” de los datos para hacer y mejorar predicciones e inferencias. 19
- membrana** Lámina de tejido orgánico, generalmente flexible y resistente. 29
- milivolt** Equivalente a una milésima de voltio. 31
- miocarditis** Inflamación del músculo cardíaco (miocardio).. 16
- mortalidad** Cantidad de personas que mueren en relación con el total de la población.. 16
- multifractal** Consiste de muchos fractales.. 18
- ondeleta** Función matemática que permite seccionar los datos en pequeños componentes en escala frecuencia–tiempo y analizarlos cada uno por separado. 16
- precursor** Conjunto de características y patrones subyacentes en los datos cuya tendencia a un comportamiento en específico puede conducir a un estado posterior o influir en su desarrollo.. 14
- prognosis** Conocimiento anticipado de algún suceso, especialmente del tiempo meteorológico.. 16
- protocolo** Conjunto de reglas que, ya sea por norma o por costumbre, se establecen para actos oficiales o solemnes, ceremonias y otros eventos.. 16
- septal** Del séptum o relacionado con él.. 30
- sesgo** Valor de error que mide la diferencia entre la esperanza matemática y el valor real de la variable. 17
- sístole** Representa la contracción muscular que permite su eyección. 26
- súbita** Que se produce de pronto, sin preparación o aviso.. 18
- ultrasonido** Vibración mecánica de frecuencia superior a la de las que puede percibir el oído.. 16
- umbral** Parte inicial o primera de un proceso o actividad.. 29
- ventricular** Se refiere a cada una de las cámaras inferiores del corazón. 18

Índice general

0.1. Resumen	14
0.2. Abstract	15
1. Introducción	16
1.1. Estado del Arte	18
2. Objetivos	20
2.1. Objetivo General	20
2.2. Objetivos específicos	20
3. Planteamiento del Problema.	21
4. Justificación de la Investigación	22
5. Marco Teórico	23
5.1. Terminología y Precisiones	23
5.1.1. Definición de Precursor y Precursores fisiológicos	23
5.1.2. Terminología Importada	23
5.2. Corazón y fisiología cardíaca	24
5.3. Ciclo cardíaco	25
5.4. Mecánica Auricular y ventricular	26
5.4.1. Función de bomba de las aurículas	26
5.4.2. Función de bomba de los ventrículos	26
5.4.2.1. Llenado de ventrículos durante la diástole	26
5.4.3. Vaciamiento de los ventrículos durante la sístole	26
5.4.3.1. Periodo de contracción isovolumétrica	26
5.4.4. Periodo de eyeción	26
5.4.4.1. Periodo de relajación isométrica (isovolumétrico)	27
5.5. Función de las válvulas cardíacas	27
5.5.1. Válvulas auriculoventrículares (AV)	27
5.5.2. Músculos palpitares	27
5.6. Excitación rítmica del corazón	28
5.6.1. Sistema de conducción y excitación	28
5.6.1.1. Nódulo sinoauricular	29
5.6.1.2. Ritmicidad automática	29
5.6.1.3. Vías internodales y transmisión del impulso a las aurículas	29
5.6.1.4. Nodulo auriculoventricular	30
5.6.1.5. Sistema de Purkinje	30
5.6.1.6. Transmisión a los ventrículos	30
5.7. Electrocardiograma	31
5.7.0.1. Voltaje y calibración cronológica	31
5.8. Derivaciones electrocardiográficas	32
5.8.1. Derivaciones bipolares de las extremidades	32
5.8.2. Triángulo de Einthoven	32

5.8.2.1. Ley de Einthoven	32
5.8.3. Electrocardiograma normal en las 3 derivaciones bipolares	33
5.9. Derivaciones precordiales	33
5.9.1. Derivaciones unipolares	33
5.10. Interpretación electrocardiográfica	33
5.11. Enfermedades Cardiovasculares (ECV):	34
5.11.1. Infarto agudo de miocardio (IAM)	34
5.11.2. Fibrilación Atrial (FA):	34
5.11.3. Insuficiencia cardíaca congestiva (ICC):	35
5.12. Variabilidad del ritmo cardíaco (HRV)	36
5.12.1. Definición de la HRV	36
5.12.1.1. Reflejos de control en la HRV	36
5.13. Técnicas de análisis lineal	37
5.13.1. Momentos Estadísticos	37
5.13.2. Media	37
5.13.3. Varianza	37
5.13.4. Coeficiente de Variación	38
5.13.5. Skewness o Sesgo	38
5.13.6. Curtosis	38
5.13.7. Señal y Serie de Tiempo	39
5.13.8. Ventanas y Técnica de Ventaneo	39
5.14. Técnicas de análisis No-Lineal	40
5.14.1. Sistema Complejo/s y Complejidad:	40
5.14.2. Información y entropía de la información	40
5.14.3. Entropía Aproximada	41
5.14.4. Entropía Muestral	41
5.14.5. Entropía Espectral	42
5.14.6. Objetos fractales y procesos auto-similares	42
5.14.7. Dimensión Fractal de Higuchi	44
5.14.8. Exponente de Hurst	45
5.14.9. Análisis fractal y series de tiempo con procesos auto-similares	45
5.14.9.1. Ruido Rosa ó Ruido $1/f$	45
5.14.10. Análisis de fluctuación sin tendencia (DFA)	46
5.14.11. Análisis por Gráficos de Poincaré	47
5.15. Procesamiento de señales con wavelets	48
5.15.1. Wavelets	48
5.15.2. La Transformada Wavelet	48
5.15.3. Densidad espectral de potencia de ondeletas	49
5.16. Clasificación por medio de Machine Learning	50
5.16.1. El paradigma de Machine Learning	50
5.16.1.1. Algoritmo	50
5.16.1.2. Algoritmos Supervisados y No Supervisados	50
5.16.1.3. Modelo de Aprendizaje	51
5.16.1.4. Dataframe	51
5.16.1.5. Instancia	51

5.16.1.6. Característica	51
5.16.1.7. Objetivo	51
5.16.1.8. Tarea	51
5.16.1.9. Predicción	51
5.16.2. Clasificadores Binarios	52
5.16.3. Evaluación de Clasificadores Binarios	52
5.16.4. Matríg de Confusión	53
5.16.4.1. Exactitud	53
5.16.4.2. La Precisión	53
5.16.4.3. Sesgo	54
5.16.4.4. Sensibilidad y Especificidad	54
5.16.5. Curvas ROC y curvas de recuperación-precisión	54
5.17. Arboles de Decisión	55
5.17.1. Extremely Randomized Trees Classifier	55
5.18. Análisis de Componentes Principales	56
5.18.1. Hiperplano	56
6. Desarrollo experimental	57
6.1. Propuesta de solución	57
6.1.0.1. Ingeniería de Datos	57
6.1.0.2. Procesamiento y análisis de señales	57
6.1.0.3. Clasificación	58
6.2. Ingeniería de Datos	58
6.2.1. DUMMY	58
6.2.1.1. Bases de datos consideradas en DUMMY	59
6.2.2. MIMIC-III	60
6.2.2.1. Lista de Tablas	60
6.2.2.2. MIMIC-III Waveform Database Matched Subset	62
6.2.3. Descarga de bases de datos	65
6.2.4. Lectura de Datos y estructuración de Datasets	66
6.2.4.1. Reducción de las bases	66
6.2.4.2. Preparación del Dataset para el pre-procesamiento	66
6.3. Procesamiento y análisis de señales	69
6.3.1. Pre-procesamiento de la señal ECG	69
6.3.1.1. Lectura de señales ECG	69
6.3.1.2. Reducción de ruido en la señal	70
6.3.1.3. Corrección de Línea base errante	70
6.3.1.4. Detección de picos y obtención de señal R-R	71
6.3.2. Metodología para el procesamiento de la señal R-R	73
6.3.2.1. Técnica de ventaneo dinámico	73
6.3.3. Análisis Lineal de la señal R-R	74
6.3.3.1. Visualización de momentos estadísticos	75
6.3.3.2. Generación de archivos csv	78
6.3.3.3. Lectura y limpieza de datos	78
6.3.3.4. Análisis de distribución de las métricas lineales	78

6.3.3.5.	Gráficos de dispersión de los momentos estadísticos	80
6.3.3.6.	Comparativa Grupo sano vs ECV	81
6.3.4.	Análisis No Lineal	83
6.3.4.1.	Análisis sobre fractalidad y no-linealidad de las señales	84
6.3.4.2.	Análisis de Poincaré	85
6.3.4.3.	Análisis por Transformada Wavelet	85
6.3.4.4.	Inspección por medio del Espectro de Potencia de Wavelet	86
6.3.4.5.	Generación de archivos csv	87
6.3.4.6.	Lectura y limpieza de datos	87
6.3.4.7.	Gráficos de dispersión	88
6.3.4.8.	Análisis no lineal (Grupo sano vs ECV)	89
6.3.5.	Clasificación	90
6.3.5.1.	Reducción dimensional por PCA	90
6.3.5.2.	Modelo de clasificación supervisado	90
6.3.5.3.	Análisis no lineal para patologías cardíacas (FA,CC e IM)	92
6.3.5.4.	Gráficas de dispersión	93
6.3.5.5.	Reducción dimensional por PCA Entre patologías	94
6.3.5.6.	Clasificación supervisada	95
6.3.5.7.	Comparación entre FA e ICC	96
6.3.5.8.	Comparación entre FA e IM	97
6.3.5.9.	Comparación entre ICC e IM	98
7.	Análisis e Interpretación de Resultados	100
7.1.	Análisis Lineal	100
7.2.	Análisis no Lineal	100
7.3.	Modelo de clasificación	101
8.	Discusión	102
9.	Conclusiones	104
10.	Recomendaciones y/o Sugerencias para investigaciones futuras	105

Índice de figuras

5.1.	Anatomía del corazón y flujo sanguíneo a través de las cavidades y válvulas cardíacas.	24
5.2.	Eventos ocurridos en el ciclo cardíaco por secuencia temporal.	25
5.3.	Válvulas mitral y aórtica junto con músculos papilares.	28
5.4.	Nódulo del seno,sistema de Purkinje del corazón, nódulo auricoventricular(AV), vías inter-nodales y auriculares.	29
5.5.	Electrocardiograma normal con la ondas P, T y el complejo QRS.	31
5.6.	El triángulo de Einthoven con las derivaciones bipolares I,II,III y unipolares aVR, aVI, aVF. .	32
5.7.	Ventaneo de una Señal	39
5.8.	Helecho de Barnsley: Objeto fractal generado por computadora	43
5.9.	Auto-similitud de una Serie de Tiempo	44
5.10.	Gráfico de Poincaré de datos aleatorios	47
5.11.	Ondeleta de Morlet	48
5.12.	WPS de serie de tiempo relacionada con el índice HSI	49
5.13.	Matríz de confusión simple	53
5.14.	Curva ROC de un regresión logística	54
5.15.	Curva RC de un regresión logística	55
5.16.	Ejemplo de la reducción dimensional por PCA de 3 características a 2 componentes prin-cipales	56
6.1.	Descripción general de la base de datos de cuidados intensivos MIMIC-III.	62
6.2.	Distribución de género de los registros extraídos de MIMIC-III	63
6.3.	Distribución de grupos étnicos de los registros extraídos de MIMIC-III	64
6.4.	Diagrama de Flujo para descarga de MIMIC III	65
6.5.	Diagrama de flujo del pre-procesamiento para bases de prueba	67
6.6.	Diagrama de flujo del Pre-procesamiento para MIMIC	68
6.7.	Señal con anotaciones correctas(abajo) e incorrectas(arriba)	69
6.8.	Señal antes(Arriba) y después (abajo) del filtro.	70
6.9.	Media móvil generada(superior) y Señal con el ajuste de línea(inferior)	71
6.10.	Detección de picos y señal de variabilidad cardíaca obtenida.	72
6.11.	Diagrama de flujo del análisis lineal	74
6.12.	Gráfica de momentos Sano	75
6.13.	Gráfica de momentos Sano	76
6.14.	Gráfica de métricas lineales de FA	77
6.15.	Gráfica de métricas lineales de ICC	77
6.16.	Distribución del primer comentario en casos saludables	79
6.17.	Distribución del primer comentario en casos FA	79
6.18.	Gráfico de dispersión de la media VS varianza en los 2 grupos de estudio.	80
6.19.	Gráfico de dispersión de la media VS sesgo en los 2 grupos de estudio.	80
6.20.	Gráfico de dispersión de la media VS curtosis en los 2 grupos de estudio.	80
6.21.	Gráfico de dispersión de la varianza VS sesgo en los 2 grupos de estudio.	80
6.22.	Histograma de la media en grupo sano y con ECV	81
6.23.	Diagrama de flujo del análisis no lineal	83
6.24.	Gráfica de métricas no-lineales de FA	84

6.25. Gráfica de métricas no-lineales de ICC	84
6.26. Gráfica de métricas no-lineales casos Sanos	84
6.27. Gráfica de cociente R	84
6.28. Gráfico de Poincaré para una señal saludable	85
6.29. Transformada wavelet para registro sano de R-R	85
6.30. WPS de registro de AF	86
6.31. WPS de registro de ICC	86
6.32. WPS de registro Sanos	86
6.33. Gráfico de dispersión del exponente de Hurst vs DFA.	88
6.34. Gráfico de dispersión DFA vs entropía aproximada.	88
6.35. Gráfico de dispersión DFA vs PSD.	88
6.36. Gráfico de dispersión Hurst vs entropía aproximada.	88
6.37. Histogramas de la dimensión fractal calculada por DFA en grupo sano y con ECV	89
6.38. Uso de PCA a partir de CSV	90
6.39. Matriz de confusión obtenida del proceso de prueba del algoritmo de clasificación entrenado con un tamaño de prueba de n =150	90
6.40. Curva ROC del algoritmo de clasificación implementado para diferenciar sanos vs ECV.	91
6.41. Curva R-C del algoritmo de clasificación implementado para diferenciar sanos vs ECV.	91
6.42. Histograma mostrando la técnica de dimensión fractal por DFA para las 3 patologías	92
6.43. Histograma mostrando la técnica de entropía muestral para las 3 patologías	93
6.44. Gráfico de dispersión del exponente de Hurst vs el índice fractal por DFA.	93
6.45. Gráfico de dispersión del índice fractal por DFA vs entropía muestral.	93
6.46. Gráfico de dispersión del índice fractal por DFA vs Espectro de Potencia.	94
6.47. Gráfico de dispersión del espectro de potencia vs entropía muestral.	94
6.48. Gráfica de dispersión para las 3 componentes principales obtenidas por el PCA	94
6.49. Matriz de confusión resultante del modelo de clasificación para las 3 patologías con una muestra de entrenamiento del 20%	95
6.50. Relación entre el número de componentes principales analizadas y la eficiencia del modelo	96
6.51. Matriz de confusión resultante de la clasificación del algoritmo de FA e ICC	96
6.52. Curva ROC para el modelo de clasificación entrenado con los datos de registros de FA e ICC	97
6.53. Curva R-C para el modelo de clasificación entrenado con los datos de registros de FA e ICC.	97
6.54. Matriz de confusión resultante de la clasificación del algoritmo de FA e IM	97
6.55. Curva ROC para el modelo de clasificación entrenado con los datos de registros de FA e IM	98
6.56. Curva R-C para el modelo de clasificación entrenado con los datos de registros de FA e IM.	98
6.57. Matriz de confusión resultante de la clasificación del algoritmo de ICC e IM	98
6.58. Curva ROC para el modelo de clasificación entrenado con los datos de registros de IM e ICC	99
6.59. Curva R-C para el modelo de clasificación entrenado con los datos de registros de IM e ICC	99
1. Histograma de varianzas casos sanos y ECV.	109
2. Histograma de Sesgo casos sanos y ECV.	109
3. Histograma de Curtosis casos sanos y ECV	109
4. Histograma de entropía aproximada para casos sanos y con ECV	110
5. Histograma de entropía muestral para casos sanos y con ECV	110
6. Histograma de exponente de Hurst para casos sanos y con ECV	110
7. Histograma de dimension fractal por Higuchi para casos sanos y con ECV	110

8.	Histograma del espectro de potencia para casos sanos y con ECV	110
9.	Histograma de entropía aproximada para las 3 patologías	111
10.	Histograma de dimensión fractal por Higuchi para las 3 patologías	111
11.	Histograma de exponente de Hurst para las 3 patologías	111
12.	Histograma de espectro de potencia para las 3 patologías	111
13.	Gráfica de dispersión sesgo vs Curtosis	112
14.	Gráfico de dispersión varianza vs Curtosis	112
15.	Gráfica de dispersión Higuchi vs DFA	112
16.	Gráfico de dispersión Hurst vs PSD	112
17.	Gráfica de dispersión Hurst vs entropía muestral	113
18.	Gráfico de dispersión Entropía muestral vs entropía aproximada	113
19.	Gráfica de dispersión DFA vs Higuchi	113
20.	Gráfico de dispersión exponente de Hurst vs espectro de potencia	113
21.	Gráfica de dispersión exponente de Hurst vs entropía muestral	114
22.	Gráfico de dispersión entropía muestral vs aproximada	114
23.	Metricas lineales de casos de FA	115
24.	Metricas lineales de casos de ICC	116
25.	Metricas lineales de casos de Control	117
26.	Metricas no-lineales de casos de FA	118
27.	Metricas no-lineales de casos de ICC	119
28.	Metricas no-lineales de casos de Control	120

Resumen

Diversos sistemas biológicos pueden ser descritos como sistemas complejos, los cuales son integrados por un número grande de componentes interrelacionados y, a menudo, la interacción es del tipo no lineal abarcando diversas escalas espacio-temporales. En el contexto de la variabilidad cardiaca, los análisis de la actividad cardiaca en condiciones sanas y patológicas han sido descritas como un paradigma de la complejidad en diferentes investigaciones durante las últimas décadas. Debido a las transiciones entre estados o variabilidad en un sistema complejo, surge la necesidad de identificar y describir los mecanismos que actúan en el control del sistema, y reconocer los precursores que indiquen la inminente aparición de un evento crítico en los mismos. En este trabajo, mediante este enfoque, se pretende la caracterización de precursores asociados a patologías como el infarto al miocardio, fibrilación atrial e insuficiencia cardíaca congestiva. Nuestro procedimiento contempla analizar series de tiempo derivadas de la actividad eléctrica del corazón mediante técnicas lineales, no lineales y de sistemas complejos. Los registros provienen de bases de datos clínicos especializadas, los cuales hacen factible extraer patrones que sirvan como precursores para las patologías mencionadas. Adicionalmente, se contempla en etapas futuras la evaluación de la eficiencia en un algoritmo clasificador supervisado que busque en investigaciones futuras trabajar con datos en tiempo real para diferenciar entre estados saludables y patológicos.

Palabras clave: sistemas complejos, dinámica cardíaca, infarto al miocardio, fibrilación atrial, falla cardíaca congestiva.

Abstract

Many biological systems could be described as complex systems due to the number of interrelated elements and the non-linearity of their dynamics. The pathological cardiac dynamics recorded through electrocardiogram have been described within the paradigm of complexity in different investigations. Due to the transitions between states or variability in a complex system, the need to identify and describe the mechanisms that act in the control of the system and recognize the precursors that indicate an imminent appearance of a critical event arises. In the present work, using this approach, we pretend to characterize the precursors associated with myocardial infarction, atrial fibrillation and congestive heart failure. Our procedure is based on electrical heart time series analyses through nonlinear techniques and complex systems related. The registers come from specialized clinical data bases, which makes feasible pattern extraction that can be identified as precursors for the mentioned pathologies. Additionally, evaluation of the efficiency for a supervised classifier algorithm is discussed, expecting to be useful in future investigations with real time data to differentiate between pathological and healthy states.

Keywords: complex systems, cardiac dynamics, myocardial infarction, atrial fibrillation, congestive heart failure.

Capítulo 1

Introducción

El análisis de las variables fisiológicas es un área de interés tanto clínico como académico, por su relevancia para identificar y diagnosticar de forma no invasiva la aparición de patologías a menudo relacionadas con alteraciones espacio-temporales. En este sentido, la variabilidad cardíaca, registrada a través del electrocardiograma (ECG), permite evaluar la actividad eléctrica del corazón y es una herramienta ampliamente utilizada para el diagnóstico de arritmias, anomalías cardíacas congénitas, miocarditis e isquemia [1].

Este conjunto de enfermedades cardiovascularcardiovasculares representan la principal causa de muerte en el mundo, siendo responsables del 30 % de las defunciones mundiales, tan solo en el año 2016 registrando alrededor de 17.1 millones de muertes [2]. Teniendo además estimaciones de la Organización Mundial de la Salud o OMS (del inglés World Health Organization o WHO) que para el 2030 casi 23,6 millones de personas morirán por alguna enfermedad cardiovascular, principalmente por cardiopatías y accidentes cerebrovasculares[3].

La prevención y un diagnóstico oportuno se vuelven indispensable para reducir la tasa de mortalidad de estas enfermedades. Durante las últimas décadas se han adecuado herramientas que buscan un diagnóstico temprano tales como biomarcadores, imágenes de tomografías y ultrasonido, sin embargo, aún se continua explorando su diferenciabilidad para distintas afectaciones cardíacas, además de requerir personal especializado en su manejo, incrementando costos de operatividad en dichas investigaciones.

El procedimiento para el diagnóstico de estas patologías actualmente incluye la lectura sistémica del ECG por parte de un médico experto de los complejos cardíacos buscando reconocer patrones de alteración en la señal como lo son segmentos de magnitud, intervalos de tiempo, nivel y características morfológicas en las ondas T,P y el complejo QRS.

Estos patrones se comparan en rangos definidos como saludables y con base en ellos se tienen criterios para evaluar la función cardíaca en el paciente. Sin embargo, las fluctuaciones de esta señal se hacen evidentes una vez que el flujo eléctrico del corazón se muestra significativamente alterado.

Señales derivadas del electrocardiograma tales como la variabilidad del ritmo cardíaco (HRV por sus siglas en inglés), han demostrado ser una herramienta confiable para la prognosis en pacientes con patología cardíaca además de proporcionar alertas tempranas de riesgo en las mismas [4].

Múltiples investigaciones se han enfocado en el reconocimiento de patrones que busquen auxiliar en el diagnóstico temprano de alteraciones cardíacas basadas en la señal HRV, en este marco, la propuesta se centra en analizar bases de datos caracterizadas de electrocardiograma para algunas de las más comunes afectaciones dentro de este grupo como lo son el Infarto Agudo al Miocardio, la Fibrilación Atrial y la Insuficiencia Cardíaca Congestiva, con la finalidad de obtener precursores de estas patologías dadas principalmente por propiedades extraibles en su señal aplicando técnicas de entropía de información, transformada de ondeletas, gráficos de Poincaré y dimensión fractal. Con la finalidad de desarrollar un algoritmo clasificador supervisado para una primera distinción entre señales saludables y no saludables, buscando ser incorporado en un protocolo de diagnóstico automatizado y confiable.

Como un primer acercamiento, se contempla la implementación de técnicas lineales (media, varianza, curtosis, sesgo) a través de un ventaneo dinámico y un preprocesamiento, para verificar la no diferenciabilidad y la separación del régimen gaussiano en las señales patológicas analizadas. Esto con el fin de sentar antecedentes para justificar el uso de técnicas no lineales en las etapas futuras de la investigación.

Debido a que, en la literatura se hace uso indistinto de los términos para identificar la variabilidad cardíaca tales como HRV, intervalo RR, e intervalos NN se hará uso indistinto de los mismos y cuando se refiera específicamente a la señal electrocardiográfica pura será identificada como ECG. Mientras que, para las patologías mencionadas, Infarto Agudo al Miocardio, Fibrilación Atrial e Insuficiencia Cardíaca Congestiva se hará referencia a ellas en el texto con las siglas IAM, FA e ICC respectivamente.

Estado del Arte

Estudios efectuados han mostrado el comportamiento no lineal de la patología cardíaca y la necesidad de implementar análisis complejo para el modelado de su comportamiento. El trabajo de Tulppo et.al. probó la hipótesis de que existe una organización fractal de la dinámica autonómica en la regulación de la frecuencia cardíaca. Este trabajo sienta las bases del proyecto planteado, abriendose a la exploración y análisis de las características caóticas del funcionamiento del sistema encargado de la regulación cardiaca [5].

Starkova y Sovoliev en su trabajo del año 2009 han utilizado multiples técnicas no lineales como la transformada de ondeleta y análisis de multifractalidad para observar la diferenciabilidad en señales de ECG de pacientes los cuales presentaron muerte súbita cardíaca, observando en el dominio de la frecuencia diferencias en cuanto al ruido $1/f$ de los grupos de control saludables y patológicos. concluyendo puede utilizarse como un marcador para diagnóstico. Es de interés en este trabajo implementar la técnica de transformada de ondeleta implementada en dicha investigación, sin embargo se pretende analizar bases con un número mayor de datos que los 24 registros que fueron utlizados en dicho trabajo [6].

Myriam Mohebbi y Hassan Ghassemian llevaron a cabo en el año 2010 la detección de fibrilación auricular paroxística por medio de un algoritmo basado en el análisis de HRV, encontrando 14 marcadores para eventos críticos, las cuales consisten en características espectrales, bi-espectrales y no lineales(energía, bandas de baja frecuencia, segmentos temporales), llegando a intervalos de eficiencia superiores al 90 %, se pretende en este trabajo la aplicación de los gráficos de Poincare, la cual fue una técnica no lineal implementada dentro de este análisis además de la obtención de los intervalos R-R. [7].

Reko Kemppainen en su tesis del año 2012, reporta la eficacia de patrones extraídos de ECG para su aplicación dentro de sistemas de predicción tales como parámetros de control (magnitud de QRS,ángulo QRS-T, posición del segmento QRS) y morfológicos (nivel de segmento S-T, asimetría de la onda T, área del segmento T) para medir sus respectivas tazas de éxito en la predicción de arritmias ventriculares, logrando una eficiencia del 93 %. Esta tesis es de relevancia para nuestra investigación, ya que se analizan mas de 14 patrones en el dominio del tiempo que presentaron un óptimo desempeño, su técnica para probar estas tazas de éxito se encuentra basada en la distancia de Mahalanobis con grupos de control saludables, concluyendo se requieren de técnicas mas sofisticadas para realizar estas pruebas como el algoritmo clasificador que se pretende implementar en nuestro trabajo además de un mayor número de registros de tiempo [8].

En el artículo publicado en el 2015, Shalini Sahay realiza una comparación de técnicas de extracción de información de ECG para diagnóstico de enfermedades cardíacas, las técnicas empleadas se basan en métodos de lógica difusa, algoritmos genéticos, redes neuronales, transformada de ondeletas y transformada Karkunen-Loven, teniendo cada una sus ventajas y limitaciones, pero siendo el análisis por ondeletas aquel que presenta criterios exactitud y predictibilidad (95 %), siendo de nuevo una técnica que se pretende implementar en nuestro estudio [9].

Yagmur Öztürk en 2019, ha estudiado la técnica de análisis de fractales como una herramienta de diagnóstico para enfermedades cardíacas con el uso de ECG, en dicho estudio se comparó la eficacia del método de Higuchi y conteo de cajas, concluyendo este último no se recomienda para su aplicación debido a la alta variación que existe en sus resultados, siendo entonces el método de Higuchi el señalado para encontrar irregularidades cardíacas, siendo una técnica que se pretende aplicar. [10]

Recientemente, Stephen Guth y Themistoklis P. Sapsis proponen un nuevo criterio para optimizar

los predictores que tiene en cuenta las características especiales de dichos eventos, en particular su baja tasa de ocurrencia. Tomando como referencia los modelos de Machine Learning propuestos en el trabajo anteriormente citado (modelo de flujo de Kolmogorov, modelo Majda – McLaughlin – Tabak (MMT), indicador bimodal), el trabajo propuesto busca comprobar la eficacia de dichos modelos que han sido probados en eventos críticos que ocurren en sistemas dinámicos complejos en las señales biológicas correspondientes a las patologías de interés [11].

Al tiempo del desarrollo de la presente investigación, trabajos como el de Lahcen El bouny, Mohammed Khalil y Abdellah Adib presentan un nuevo método de aprendizaje profundo de extremo a extremo (E2E-DL) para el diagnóstico de enfermedades cardíacas a partir de una señal de ECG de un solo canal con una precisión del 99.57 % (utilizando la técnica de validación cruzada 10-Fold Cross Validation) [12].

Análisis como el que expone el trabajo de G. N. Georgieva-Tsaneva, la variabilidad de la frecuencia cardíaca mediante el uso de métodos estadísticos de análisis que se utilizan tradicionalmente para estudiar la funcionalidad del corazón y mediante el análisis de fluctuación sin tendencia (DFA) . [13]

Capítulo 2

Objetivos

Objetivo General

Analizar eventos de Fibrilación Atrial, Infarto Agudo del Miocardio e Insuficiencia cardíaca congestiva por medio de análisis complejo de series de tiempo utilizando gráficos de Poincaré, Entropía de Información, Análisis Fractal y Transformada de Ondeleta para la caracterización de precursores fisiológicos.

Objetivos específicos

- Obtener los registros electrocardiográficos correspondientes a Infarto agudo al miocardio, fibrilación atrial e insuficiencia cardíaca congestiva en las bases de datos para su clasificación.
- Preprocesar las señales de cada registro para la obtención de la señal de variabilidad cardíaca (HRV).
- Aplicar técnicas lineales (media, curtosis, sesgo) para verificar la no diferenciabilidad y la separación respecto al régimen Gaussiano del conjunto de datos.
- Aplicar técnicas no lineales de entropía de información, transformada de ondeletas, gráficos de Poincaré y fractalidad a los registros ECG Y HRV para obtener patrones morfológicos, magnitud y de control en las ondas de ECG.
- Verificar la eficiencia del conjunto de precursores obtenidos con las técnicas no lineales mediante un algoritmo clasificador supervisado de patologías para comparar los registros del grupo de control saludable existente en la base de datos.

Capítulo 3

Planteamiento del Problema.

La evaluación del ECG por parte de un médico especialista depende de su experiencia y habilidad para reconocer patrones anómalos en las señales, mismas que pueden variar de persona en persona, basándose en un registro temporal en el cual no siempre es factible visualizar propiedades significativas.

Previamente, se han analizado con técnicas no lineales, registros continuos de ECG en grupos de control saludables, encontrando fluctuaciones críticas en un 12 % de los individuos, alejándose significativamente de lo que se esperaría de un diagnóstico saludable, relacionándose con patologías en un estado temprano, para las cuales, por medio de un análisis complejo serían factibles de identificar con propiedades fractales, de entropía y energía [14].

La primer problemática que surge, de acuerdo con la literatura, es establecer algoritmos de pre-procesamiento para extraer componentes de forma automática en el ECG como el intervalo R-R. Eliminando ruido ambiental y latidos cardíacos ectópicos sin excluir otras propiedades caóticas importantes que sean indispensables para efectuar los estudios.

Una segunda problemática a considerar, es el acceso a volúmenes caracterizados y masivos de datos de fuentes confiables, los cuales contengan registros de individuos previamente diagnosticados, a partir de las cuales se obtengan mayores muestras poblacionales que reduzcan la incertidumbre y añadan representatividad a los fenómenos estudiados con el fin de establecer métricas compatibles con mayores grupos poblacionales.

La obtención de precursores en patologías cardíacas se han reportado en la literatura desde hace décadas y ha sentado las bases para definir aquellos conjuntos con mayor probabilidad de extraer métricas confiables. Para los casos en concreto de IM, se han demostrado fluctuaciones y un comportamiento no estacionario en la disfunción ventricular con cambios en las propiedades multifractales. [15]. En el caso de FA, la arritmia más común en la población, investigaciones demuestran que el comportamiento del corazón durante esta patología sigue una dinámica no lineal y se ha logrado extraer marcadores de eventos críticos en el intervalo N-N [4]. Existen sin embargo, patologías las cuales no han sido estudiadas lo suficiente con este enfoque, tal es el caso de ICC, para la cual se han analizado bases de datos en busca de dispersiones en las ondas electrocardiográficas, sin embargo no se han obtenido características diferenciables aún [16]. En el contexto de este trabajo y para evitar posibles confusiones por ambigüedad, el uso del término “precursor” es definido de manera puntual dentro del marco teórico.

Es de interés para este trabajo, el acceso a la base de datos MIMIC-III, la cual contiene 61,532 registros clínicos de pacientes en unidades de cuidado intensivo y la cual ha sido utilizada en trabajos mencionados en el estado del arte por su información categorizada y registros de ECG para pacientes con enfermedades cardiovasculares, mediante la cual se diseñarán los algoritmos correspondientes de pre-procesamiento, extracción de la señal HRV y se realizarán pruebas con técnicas lineales con el objetivo de mostrar el comportamiento caótico y las fluctuaciones críticas en señales patológicas de ECG, para sentar las bases de un estudio mediante análisis complejo con los datos clínicos certificados de MIMIC-III.

Tomando este conjunto de enfermedades, previamente analizadas en diferentes investigaciones y obteniendo el acceso a una base de datos estructurada de las mismas, además de aplicar algoritmos de pre-procesamiento robustos y su respectivo análisis con técnicas complejas, el reto siguiente para este trabajo se vuelve en el reconocimiento de patrones que sean integrados en un algoritmo de aprendizaje supervisado con el fin de evaluar su desempeño para reconocer alteraciones en señales de prueba saludables y no saludables, logrando en un futuro integrarse en un software para un diagnóstico automatizado y temprano.

Capítulo 4

Justificación de la Investigación

Las enfermedades cardiovasculares, son la principal causa de muerte en el mundo, representando el 30 % de las defunciones mundiales según la OMS (17.1 millones de defunciones en el año 2016) [2].

Organismos internacionales como la red europea del corazón (EHN), han sugerido modelos que busquen atacar esta problemática recolectando la mayor cantidad de datos y estadísticas posibles de pacientes para describir con precisión factores de riesgo que repercutan directamente en la supervivencia de este conjunto de enfermedades [17].

En décadas recientes, los hospitales se han adecuado a un registro digital de datos clínicos, tan solo en Estados Unidos reportando un incremento del 7 % al 91 % entre 2008 y 2014 [18]. A pesar de este avance, no se ha buscado aprovechar toda esta información que represente mejoras con los servicios hospitalarios y de diagnóstico, en parte por su difícil acceso libre a investigadores debido a normas bioéticas.

En el año 2010, surgió la base de datos MIMIC-II, años posteriores actualizándose a MIMIC-III, siendo una base de datos certificada por el Massachusetts Institute of Technology (MIT) para fines de investigación y de acceso libre, teniendo que realizar un curso certificado de bioética para el manejo de esos datos que ascienden a mas de 60 mil registros de registros clínicos de pacientes internados en unidades de cuidado intensivo del Beth Israel Deaconess Medical Center entre los años 2001 y 2012, logrando una oportunidad para estudiar factores de riesgo y propiedades de la dinámica cardíaca [19].

El análisis automatizado de señales cardiovasculares ha despertado el interés de diversos equipos de investigación en los últimos 20 años, esto gracias al desarrollo de nuevas tecnologías y procesamiento computacional mediante software libre para la obtención de modelos que permitan identificar propiedades diferenciales entre conjuntos de señales patológicas.

Los resultados en dichas investigaciones no han sido concluyentes, sin embargo, han existido acercamientos que permiten inferir mientras mas datos sean analizados y recopilados, a la par que se explore con diferentes enfoques de investigación, describir precursores eficientes en apoyar un diagnóstico temprano para estas enfermedades, el cual repercuta directamente en su índice de mortalidad.

La aplicación de técnicas no lineales que surgen del modelado de sistemas complejos (gráficos de Poincaré, entropía de información, transformada de ondeletas y dimensión fractal) tienen un gran potencial en la detección de enfermedades (hasta 95 % de eficacia) con alto índice de mortalidad (v. gr. en shock séptico y síndrome de insuficiencia multiorgánica) mediante la exploración de interacciones reguladoras distribuidas masivamente en mecanismos fisiológicos correspondientes [20], siendo este el enfoque planteado por este trabajo.

Este tipo de análisis puede realizarse actualmente mediante el uso de software libre (Python, Waveform Data Base y Gracelet), el cual se busca implementar y que junto con el acceso libre a estas bases de datos, se tiene una oportunidad de diseñar, probar e implementar algoritmos sin costo alguno cuyos beneficios teóricos, aportarían una base para el desarrollo e implementación de un sistema ó software futuro que proporcione herramientas de diagnóstico para un médico, sin la necesidad de utilizar recursos como equipos costosos que incrementan los costos de operatividad y mantenimiento en otros enfoques de investigaciones clínicas.

Capítulo 5

Marco Teórico

Terminología y Precisiones

Definición de Precursor y Precursores fisiológicos

La palabra precursor (derivado del vocablo latino “*prae*cursor”) es un término que suele tomar un significado coloquial como su uso adjetivo lo indica, para hacer referencia a un evento que precede a otro, generalmente anunciándola o haciéndola posible [21]. En este sentido, el término precursor tiende a especificarse como un evento puntual ó persona que sucedió o existió antes que otro evento, especialmente si se desarrolló al punto de llegar a él o tuvo una influencia en él.

En el ámbito científico, el uso de este y otros términos como “predicción”, son tratados con mucho mayor cautela y requieren de una precisión contextual debido a que, dependiendo del área de estudio, puede tomar determinado significado como es el caso, por ejemplo, de un precursor químico el cual hace referencia a una sustancia necesaria para obtener otra diferente a partir de una reacción química. En ese mismo sentido, es necesario precisar el significado de los precursores fisiológicos no como un evento temporal aislado que desencadena a otro evento, sino como un patrón de comportamiento de una serie de tiempo asociada a una señal fisiológica (para nuestro caso, una señal electrocardiográfica) ó señales derivadas cuya tendencia sea demostrablemente influyente en un cambio de estado fisiológico. Esta precisión se hace de manera similar a como estudios en sismografía lo han hecho[22],[23]. Esto debido a que es necesario evitar cualquier mal uso del término que pueda dar una mala interpretación ó un enfoque tendencioso de la información presentada en la literatura científica.

Para el contexto de este trabajo y para evitar posibles confusiones por ambigüedad, el uso del término “precursor” hará referencia a características y patrones subyacentes en los datos cuya tendencia a un comportamiento en específico puede conducir a un estado posterior o influir en su desarrollo.

Terminología Importada

En los últimos años el uso de las tecnologías computacionales se ha popularizado a un ritmo acelerado. Se anticipa que el tamaño del mercado global de estas tecnologías aumentará notablemente junto con la aparición de nuevas soluciones impulsadas por ellas en favor del mercado financiero, la investigación científica y tecnológica durante los próximos años. Es por esta popularidad y su notable relevancia que algunos términos propios de dichas tecnologías (en su gran mayoría publicadas en el idioma inglés), son usados con mayor frecuencia desde lo más cotidiano hasta las ramas de la investigación que han encontrado avances significativos en su estado del arte a través de estas tecnologías.

A lo largo de este trabajo se hará uso de ciertos extranjerismos, definiendo cuando sea pertinente un equivalente en el idioma español para expresar la idea principal que subyace en dicho término ó concepto y posteriormente continuando con el uso del término en el idioma en el que fue acuñado.

Corazón y fisiología cardíaca

El corazón, es un órgano compuesto por dos bombas separadas, una derecha encargada de impulsar el flujo sanguíneo hacia los pulmones, y una izquierda impulsándola hacia el resto de los órganos periféricos. Cada una de estas válvulas es una bomba pulsátil de dos cavidades, compuesta por una aurícula y un ventrículo. La aurícula funciona como una cavidad de entrada hacia el ventrículo, ayudando a desplazar la sangre hacia este, siendo el ventrículo la principal fuerza para impulsar la sangre por la circulación pulmonar y periférica, tal como se muestra en la figura 5.1 [24].

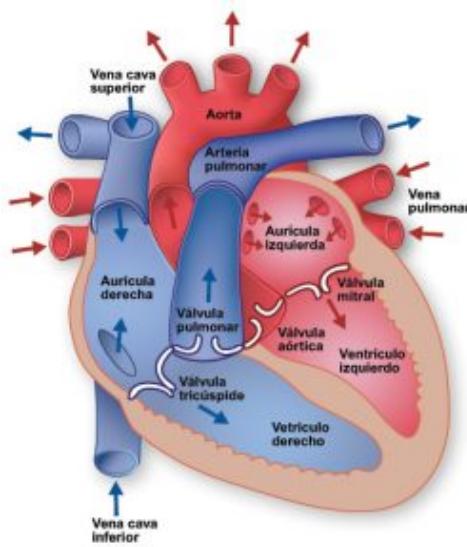


Figura 5.1: Anatomía del corazón y flujo sanguíneo a través de las cavidades y válvulas cardíacas.

La estructura logra una irrigación continua gracias a la vena cava inferior y a la vena cava superior, las cuales se conectan en la bomba derecha del corazón, ingresando la sangre desoxigenada a través de la aurícula derecha (AD) para llegar al ventrículo derecho (VD) por medio de la válvula tricúspide, una vez ahí esta es bombeada desde la arteria pulmonar principal hasta las arterias pulmonares derecha e izquierda, llegando así hasta los pulmones. En los pulmones, una vez oxigenada la sangre fluye hacia la bomba izquierda del corazón, la aurícula izquierda (AI) para transportarse desde la válvula mitral al ventrículo izquierdo (VI), el cual bombea la sangre a la aorta, la principal arteria que transporta la sangre oxigenada al resto del cuerpo.

Ciclo cardíaco

Se define como un ciclo cardíaco al intervalo temporal que abarca el comienzo de un latido cardíaco al inicio del siguiente. El inicio de un ciclo cardíaco comienza con la generación espontánea de un potencial de acción en el nódulo sinusal (NS), localizado en la pared superior de la aurícula derecha, el potencial se propaga desde el NS hacia los ventrículos a través del haz auriculoventricular (AV). El ciclo cardíaco contiene dos fases llamadas diástole y sístole, la diástole es un periodo de relajación del músculo cardíaco, en el cual la sangre ingresa a las cavidades, mientras la sístole representa la contracción muscular que permite su eyeción [24].

La frecuencia cardíaca se define como al valor inverso de la duración total del ciclo cardíaco y es una variable importante para evaluar el estrés cardíaco, si la frecuencia cardíaca aumenta, la duración del ciclo disminuye, encontrándose con una relación inversamente proporcional entre ambas variables. Al tener una frecuencia cardíaca mayor, se ve reducido en una proporción mayor la diástole, teniendo menor tiempo para un llenado de volumen completo de las aurículas y por lo tanto, menor flujo sanguíneo bombeado. Estos cambios en las variables de presión y volumen se ilustran en la Figura 5.2.

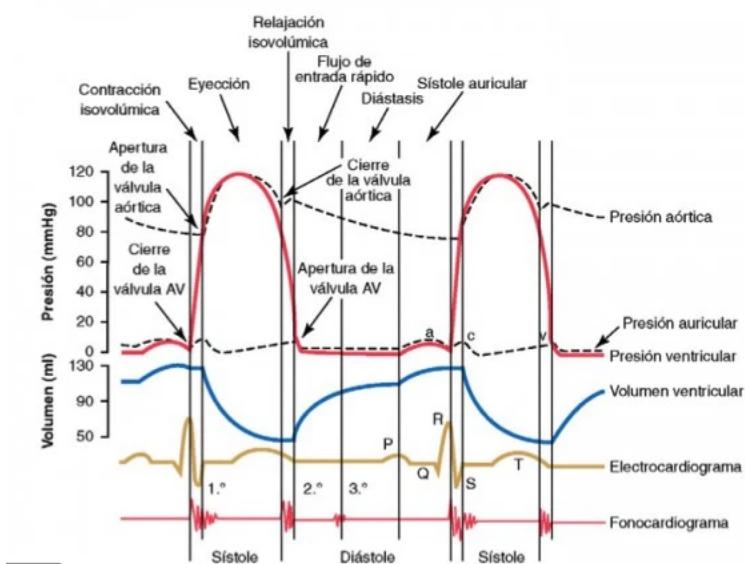


Figura 5.2: Eventos ocurridos en el ciclo cardíaco por secuencia temporal.

Mecánica Auricular y ventricular

Función de bomba de las aurículas

La sangre fluye sin interrupción desde las venas hacia las aurículas, el 75 % pasa sin transición a los ventrículos, incluso antes que las aurículas se contraigan. La contracción auricular completa es el 25 % restante de la repleción. Así, las aurículas funcionan como bombas cebadoras que aumentan la eficacia del bombeo al ventrículo. El corazón puede seguir funcionando en condiciones de reposo normal, aún sin este 25 % de eficacia, por ello aunque falten las aurículas, la persona no notará la diferencia salvo que realice ejercicio, en cuyo caso presentara signos agudos de insuficiencia cardíaca, especialmente disnea [24].

Función de bomba de los ventrículos

Llenado de ventrículos durante la diástole

En la sístole ventricular son acumulados grandes volúmenes de sangre por encontrarse cerradas las válvulas auricoventriculares (AV), tan pronto como termina la sístole, las presiones ventriculares caen a niveles diastólicos bajos y la alta presión en las aurículas permite la apertura de las válvulas AV. Esto se denomina como periodo rápido de llenado de los ventrículos. Este periodo comprende el primer tercio de la diástole, durante el tercio siguiente solo penetra una pequeña cantidad de sangre para finalmente en el último tercio las aurículas se contraigan y proporcionen un impulso adicional para introducir sangre en los ventrículos, a ello corresponde el 25 % del llenado de los ventrículos.

Vaciamiento de los ventrículos durante la sístole

Periodo de contracción isovolumétrica

Al cerrarse las válvulas AV se necesitan de 0.02 a 0.03 segundos adicionales para que el ventrículo alcance una presión capaz de abrir las válvulas semilunares (aórtica y pulmonar) contra las presiones existentes en la aorta y la arteria pulmonar. En este momento es cuando ocurre el denominado periodo de contracción isométrica, en donde un aumento de tensión aumenta en el músculo cardíaco sin producirse un acortamiento de las fibras musculares.

Periodo de eyeción

Al aumentar la presión ventricular por encima de los 80 mmHg, las presiones ventriculares logran la apertura de las válvulas semilunares. Inmediatamente comienza a expulsarse sangre de los ventrículos, de modo que aproximadamente el 70 % del vaciado de sangre se produce durante el primer tercio del

periodo de eyección, por ello se denomina como período de eyección, mientras en los dos tercios restantes se expulsa el 30 % restante, denominándose periodo de eyección lenta.

Periodo de relajación isométrica (isovolumétrico)

Al término de la sístole, comienza la relajación ventricular permitiendo que las presiones intraventriculares caigan con rapidez. Las presiones de las grandes arterias que acaban de llenarse, empujan la sangre hacia los ventrículos, lo que provoca el cierre rápido de las válvulas aórtica (VA) y pulmonar. En un intervalo de 0.03 a 0.06 segundos, el músculo cardíaco continua relajándose, disminuyéndose las presiones intraventriculares de $80mmHg$ hasta casi cero, regresando a sus valores diastólicos para la apertura de las válvulas AV y comenzar un nuevo ciclo ventricular.

Función de las válvulas cardíacas

El corazón cuenta con cuatro válvulas que controlan el flujo de la sangre en el cuerpo. La válvula tricúspide se encarga de conectar la aurícula derecha con el ventrículo derecho. La válvula pulmonar controla el flujo entre el ventrículo derecho y las arterias pulmonares, quienes se encargan de llevar la sangre a los pulmones para oxigenarse. La válvula mitral permite que la sangre ya oxigenada pase a través de la auricula izquierda para llegar al ventrículo izquierdo. Por último, la válvula aórtica, deja que la sangre oxigenada pase del ventrículo izquierdo a la aorta, donde será transportada al resto del organismo [24].

Válvulas auriculoventrículares (AV)

Las válvulas AV (tricúspide y mitral) evitan el flujo retrógrado sanguíneo desde los ventrículos a las aurículas durante la sístole. Las válvulas semilunares (aórtica y pulmonar) impiden el flujo retrógrado de sangre desde la aorta y la pulmonar hacia los ventrículos durante la diástole. Estas válvulas se cierran cuando un gradiente de presión en sentido retrógrado las empuja hacia atrás y se abren cuando un gradiente de presión anterogrado impulsa a la sangre hacia delante. Al ser constituidas por finas y delgadas capas, las válvulas auricuventrículares, no necesitan flujo retrógrado para cerrarse, en tanto que las válvulas semilunares, mucho más fuertes, requieren de un flujo potente durante unas milésimas de segundo.

Músculos palpilares

Los músculos palpilares se contraen cuando las paredes ventriculares lo hacen, estos no ayudan a cerrar las válvulas, si no que tiran de las hojuelas de las válvulas en dirección de los ventrículos, impidiendo que puedan hacer prominencia excesiva hacia las aurículas durante la contracción ventricular. Si es rota una cuerda tendinosa o si uno de estos músculos se paraliza, la válvula hará intensa prominencia a la aurícula, permitiendo un escape considerable de sangre y originar una enfermedad cardíaca grave. La figura 3 ilustra estos músculos papilares, junto con las válvulas auricoventrículares.

Válvulas aórtica y pulmonar

Existen diferencias en el modo de operación entre las válvulas aórtica y pulmonar respecto a las válvulas auricoventrículares. La presión elevada de las arterias al final de la sístole hace que las válvulas

semilunares ciernen bruscamente, en comparación con la oclusión blanda de las auriculoventriculares. En segundo lugar, la velocidad de eyeción de sangre a través de las válvulas aórtica y pulmonar es mucho mayor que a través de las válvulas auriculoventriculares, mucho más de amplias. Es por esto, que los bordes de las válvulas semilunares están sometidos a un desgaste mecánico mayor, visualizando en la figura 5.3 resistir un traumatismo mayor.

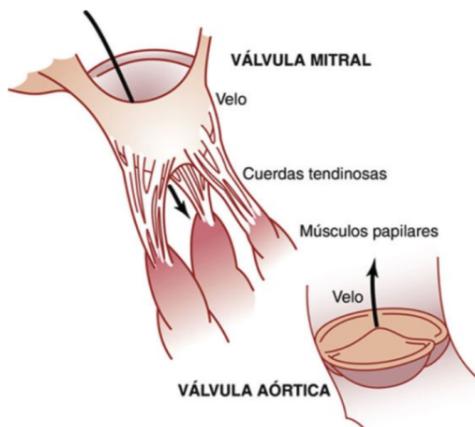


Figura 5.3: Válvulas mitral y aórtica junto con músculos papilares.

Excitación rítmica del corazón

El corazón posee un sistema especial para generar impulsos periódicos que produzcan la contracción de músculo cardíaco y para conducir estos impulsos a través de todo el corazón. Cuando este sistema funciona adecuadamente, las aurículas se contraen un sexto de segundo antes de la contracción ventricular, lo que permite el adecuado llenado de los ventrículos. Por desgracia, este sistema es muy susceptible a lesión por enfermedades cardíacas, mayormente por isquemias coronarias, que producen ritmos anómalos y escalamientos que alteran la contracción de las cavidades cardíacas afectando de manera grave la eficacia del corazón como bomba, incluso hasta el grado de producir la muerte [25].

Sistema de conducción y excitación

En la figura 5.4, se ilustran las diferentes partes que conforman este sistema. Se muestran : A) nódulo del seno, el cual genera el impulso rítmico normal autoexcitador; B) las vías internodales, que conducen el impulso desde el nódulo S-A al nódulo A—V; C) El nódulo A-V (auricoventricular), en el cual se genera el retraso antes de pasar al ventrículo; D) el haz A-V, que conduce el impulso desde las aurículas a los ventrículos y, E) los haces derecho e izquierdo de las fibras de Purkinje que conducen el impulso a los ventrículos.

Nódulo sinoauricular

El nódulo sinoauricular, también llamado nódulo del seno, es una cinta pequeña aplanada y elíptica de músculo especializado que mide aproximadamente 3mm de ancho, 15mm de largo y 1 mm de espesor. Se encuentra localizado en la parte anterosuperior de la aurícula derecha. Las fibras de este nódulo solo tienen de 3 a 5 micras de diámetro, sin embargo, estas se continúan con las fibras auriculares, extendiendo el potencial de acción que comience en ellas a las aurículas.

Ritmidad automática

Las fibras cardíacas son capaces de autoexcitarse, proceso que puede provocar una contracción rítmica automática. Esto se hace más evidente en las fibras especializadas como sistema conductor, encontrándose la mayor proporción de estas fibras en el nódulo sinoauricular. Los iones de sodio, que producen el potencial de acción tienden a difundir naturalmente hacia el interior de sus fibras a través de múltiples canales de membrana, por tanto el potencial en reposo va aumentándose entre cada dos latidos cardíacos, cuando alcanza el voltaje umbral de unos 40 milivoltios, los canales de sodio y calcio se activan, lo cual conduce a una entrada muy rápida tanto de iones de calcio como de sodio, produciendo entonces el potencial de acción. Es entonces esta permisividad de las fibras a los iones sodio la que permite la autoexcitación.

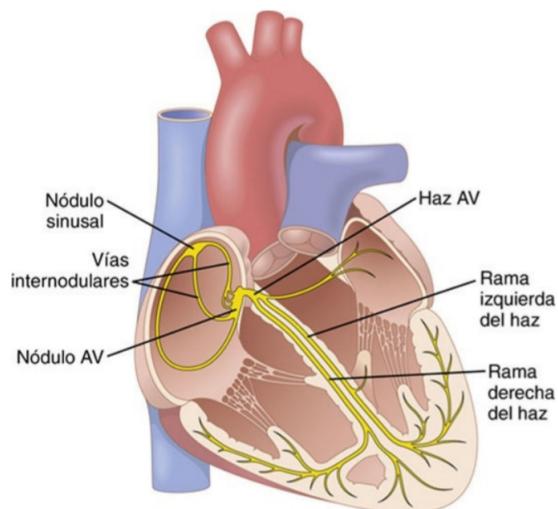


Figura 5.4: Nódulo del seno, sistema de Purkinje del corazón, nódulo auriculoventricular(AV), vías internodales y auriculares.

Vías internodales y transmisión del impulso a las aurículas

Las fibras nodales S-A se fusionan con las fibras musculares auriculares en los extremos, el potencial de acción entonces originado se difunde por el tejido muscular auricular y finalmente también llega al nódulo auriculoventricular. Esta conducción es más rápida en algunos haces de fibras musculares, la causa de que la velocidad sea más rápida es la presencia de fibras automatizadas de conducción mezcladas con el músculo auricular.

Nodulo auriculoventricular

El sistema de conducción está organizado de tal manera que el impulso cardíaco no viaja de las aurículas a los ventrículos inmediatamente, lo cual permite que las aurículas logren conducir el flujo sanguíneo hacia los ventrículos antes que empiece la contracción de estos, este retraso lo permite el nódulo auriculoventricular (AV) y sus fibras de conducción asociada. El nódulo AV se localiza en la parte septal de la aurícula derecha, detrás de la válvula tricúspide. El retraso total en el nódulo AV es de alrededor de 0.13 segundos, una cuarta parte de este intervalo se origina en las fibras de transición que son fibras más pequeñas que conectan las fibras de las vías internodales auriculares con el nódulo AV.

Sistema de Purkinje

Las fibras de Purkinje salen del nódulo AV y siguen el haz AV hacia los ventrículos y transmiten potenciales de acción a una velocidad de 1.5 a 4 metros por segundo, la transmisión muy rápida de potenciales corresponde al número elevado de enlaces entre las células cardíacas. Las fibras después de nacer en el haz AV, forman el haz AV que sigue las válvulas del corazón y desde allí penetran en el tabique ventricular. El haz auriculoventricular se divide casi inmediatamente en las ramas derecha e izquierda, situados por debajo del endocardio en los respectivos lados del tabique. Cada una de estas ramas se extiende hacia abajo en dirección del vértice del ventrículo correspondiente para dividirse en ramas pequeñas y difundir alrededor de cada cavidad ventricular.

Transmisión a los ventrículos

Una vez que el impulso ha alcanzado los extremos de las fibras de Purkinje, es transmitido a la masa muscular del ventrículo por las propias fibras musculares ventriculares, este músculo envuelve al corazón formando una doble espiral con tabiques fibrosos entre sus capas, por lo que impulso no se propaga hacia la superficie del corazón, si no que se angula hacia la superficie siguiendo la dirección de las espirales.

Electrocardiograma

EL electrocardiograma normal esta formado por una onda P, un complejo QRS y una onda T. El complejo QRS incluye 3 ondas separadas, Q, R, y S. La onda P es causada por corrientes eléctricas generadas cuando las aurículas se despolarizan , el complejo QRS está producido por corrientes nacidas cuando los ventrículos se despolarizan, mientras la onda T es causada cuando los ventrículos se recuperan del estado de despolarización. Este proceso ocurre en el músculo ventricular 0.25 a 0.35 segundos después de la despolarización y esta onda se conoce como onda de repolarización.

La onda P ocurre inmediatamente antes del principio de la contracción de las aurículas, y la onda QRS tiene lugar inmediatamente antes del principio de la contracción de los ventrículos. La onda de repolarización ventricular es la onda T del electrocardiograma normal, el músculo ventricular empieza a repolarizarse aproximadamente 0.20 segundos después de iniciarse la onda de despolarización, y completa su repolarización en todas las fibras 0.35 segundos después de comenzada la despolarización. La onda de repolarización auricular, suele quedar totalmente enmascarada por la onda QRS, por esta razón, rara vez se observa una onda T auricular [26]. La figura 5.5 ilustra las ondas de dos ciclos cardíacos consecutivos

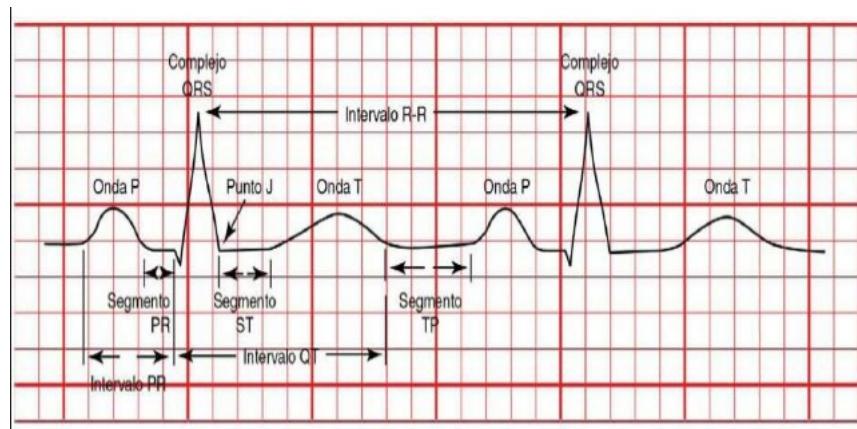


Figura 5.5: Electrocardiograma normal con la ondas P, T y el complejo QRS.

Voltaje y calibración cronológica

Todos los registros del electrocardiograma se efectúan con líneas de calibración, las líneas de calibración horizontal representan el equivalente a un milivolt (mV), con positividad en dirección ascendente y negatividad hacia abajo. Las líneas verticales del electrocardiograma son líneas de calibración en el tiempo, los voltajes de las ondas del electrocardiograma dependen de la manera en que se hayan aplicado los electrodos. Cuando los electrocardiogramas se registran con electrodos colocados en los dos brazos, o un brazo y una pierna, el voltaje del complejo QRS suele ser de un milivolt desde el vértice de la onda R hasta el fondo de la onda S, el voltaje de la onda P varía entre 0.1 y 0.3 milivolts, y el de la onda T, entre 0.2 y 0.3 milivolts.

Derivaciones electrocardiográficas

Derivaciones bipolares de las extremidades

El término bipolar hace referencia a los dos electrodos específicos localizados en el cuerpo para el registro del electrocardiograma, en este caso de las extremidades del cuerpo. Existen 3 casos para este esquema, la derivación 1 contempla la conexión de la terminal negativa del electrocardiográfo al brazo derecho y la positiva al brazo izquierdo, cuando el punto donde el brazo derecho se une al tórax es negativo respecto al punto del brazo izquierdo. En la derivación 2, la terminal negativa es unida al brazo izquierdo y la terminal positiva a la pierna izquierda. Para registrar la derivación 3, la terminal negativa es conectada al brazo izquierdo, mientras la positiva a la pierna izquierda.

Triángulo de Einthoven

El triángulo de Einthoven es un medio esquemático para señalar que los dos brazos y la pierna izquierda forman los vértices de un triángulo que rodea el corazón. Los ángulos de la parte alta representan los puntos en dónde se conectan los electrodos en los brazos, el ángulo inferior es el punto donde la punta izquierda se conecta a ambos brazos. La figura 5.6 ilustra estas conexiones de las derivaciones junto con su posición respecto a la posición del corazón.

Ley de Einthoven

La ley de Einthoven enuncia que si son registrados los potenciales eléctricos de dos de las 3 derivaciones electrocardiográficas convencionales, la tercera puede deducirse matemáticamente de las dos primeras simplemente sumándolas. Por lo que la suma de las derivaciones 1 y 3, equivalen al voltaje de la derivación 2, matemáticamente, esta ley es cierta para todo momento mientras se va registrando el electrocardiograma.

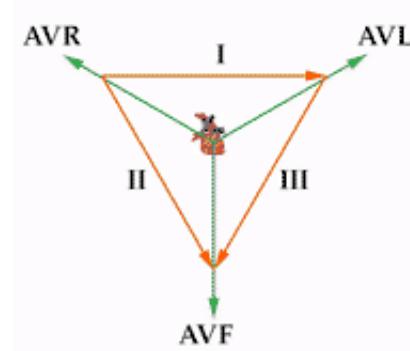


Figura 5.6: El triángulo de Einthoven con las derivaciones bipolares I,II,III y unipolares aVR, aVI, aVf.

Electrocardiograma normal en las 3 derivaciones bipolares

En estas 3 derivaciones los electrocardiogramas son muy parecidos entre si, ya que todos registran ondas P y ondas T positivas y la mayor parte del complejo QRS es positivo en cada uno. Como todos los electrocardiogramas normales son muy parecidos entre si, no tiene gran importancia cual será la derivación registrada cuando se desea detectar arritmias cardíacas, debido a que el diagnóstico de estas depende de las relaciones cronológicas, mientras que cuando se desea determinar el grado y tipo de lesión producida en ventrículos o aurículas, tiene gran importancia cuáles derivaciones se emplean, ya que estas anomalías cambian el trazado electrocardiográfico en algunas derivaciones.

Derivaciones precordiales

Estas se registran con un electrodo colocado en la cara anterior del tórax, este electrodo es conectado a la terminal positiva del electrocardiógrafo, y el electrodo negativo, denominado electrodo indiferente, se conecta a través de resistencias eléctricas del brazo derecho, el izquierdo y la pierna izquierda. En general, se registran 6 derivaciones torácicas convencionales a nivel de la pared anterior del tórax, colocando el electrodo precordial en 6 puntos los cuales se denominan derivaciones V₁,V₂,V₃,V₄,V₅ Y V₆. Como la superficie del corazón se encuentra cercana a esta pared torácica, se registra todo el potencial eléctrico del músculo cardíaco inmediatamente por debajo del electrodo, en consecuencia anomalías relativamente pequeñas de los ventrículos suelen inducir cambios intensos en los electrocardiogramas.

Derivaciones unipolares

En este tipo de registro, son conectados dos de las extremidades mediante resistencias eléctricas a la terminal negativa del electrocardiógrafo, mientras que la tercera extremidad se conecta a la terminal positiva. Cuando esta última esta colocada en el brazo derecho , esta es denominada como aVR, cuando es en el brazo izquierdo, derivación aVR y cuando en la pierna izquierda, derivación aVF.

Interpretación electrocardiográfica

Las corrientes del corazón van en una dirección particular según los momentos del ciclo cardíaco. Corrientes eléctricas fluyen del tabique y las paredes endocardíacas laterales hacia la parte externa del corazón, igualmente desde dentro de las cavidades cardíacas, directamente desde las zonas depolarizadas hacia las polarizadas. El vector sumando de las corrientes en este instante, se llama vector resultante y se obtiene a partir del centro de los ventrículos desde la base del corazón hasta la punta del mismo. En un corazón normal, la dirección media del vector durante la difusión de la onda de despolarización, llamada QRS promedio es de aproximadamente 59° [27].

Enfermedades Cardiovasculares (ECV):

Las enfermedades cardiovasculares son todas aquellas que involucran al corazón o a los vasos sanguíneos, la enfermedad de arterias , éstas incluyen:

Infarto agudo de miocardio (IAM)

El infarto agudo de miocardio, también llamado coloquialmente como paro cardíaco, la usual causa de ésta patología suele ser la enfermedad de arterias coronarias, cuando existe un riego sanguíneo insuficiente. El término herida al miocardio se usa cuando existe evidencia de valores elevados de troponina cardíaca (cTn) con por lo menos un valor arriba del percentil 99 %, se considera aguda si los valores de cTn tienen una elevación o caída. A su vez, se utiliza infarto de miocardio cuando existen heridas agudas, con evidencia clínica de isquemia aguda miocárdica, además de las variaciones de cTn y al menos una de las siguientes:

- Síntomas de isquemia miocardial.
- Cambios isquémicos en el ECG.
- Desarrollo de ondas Q patológicas-

Utilizando el ECG, no es posible inicialmente distinguir entre la manifestación de heridas al miocardio agudas o crónicas, y de isquemia aguda de miocardio [5].

A lo largo de este trabajo se utilizarán los acrónimos IM e IAM de manera intercambiable para referirse a esta patología.

Fibrilación Atrial (FA):

La fibrilación atrial es una arritmia, es decir un problema con el ritmo ó la frecuencia cardíaca que no está justificado fisiológicamente. Avances en el campo han mejorado el entendimiento que se tiene de los mecanismos electrofisiológicos tras el desarrollo de una variedad de arritmias cardíacas, son generalmente divididas en dos categorías mayores como teoría para su ocurrencia:

- Formación de pulsos abnormal o aumentada.
- Alteraciones de la Conducción (ej. reentrada de pulsos de estimulación)

La arritmia más común es la fibrilación atrial, ésta se presenta aproximadamente en el 2 % de la población en general, y 15 % de los adultos mayores. A pesar de ello aún no se acaban de conocer del todo los mecanismos que la detonan o mantienen pero sí se han desarrollado teorías y modelos tal y como se ve en el cuadro anterior, sin embargo las notables controversias entre ellas provocan que se realicen diagnósticos y tratamientos terapéuticos infectivos a éstas alturas. Esto en contraste con la taquicardia

nodal de fibrilación ventricular usando una vía accesoria, la cual tiene mecanismos bien definidos, y circuitos que pueden ser seguramente dirigidos y tienen altos índices de cura. En los últimos años se han investigado intensamente el procesamiento de señales tanto al electrocardiograma superficial (ECG), como al electrograma intracardíaco (EGM). Con base en los resultados de los registros se ha tomado ventaja del hecho en que el corazón, durante un evento de fibrilación atrial, se comporta de una manera inherentemente compleja, heterogénea y no-estacionaria, y por lo tanto la realización de análisis no lineales ha logrado llegar a avances relevantes[28][29][30].

Insuficiencia cardíaca congestiva (ICC):

La insuficiencia cardíaca congestiva es un síndrome que puede ser causado por una variedad de anomalías, que incluyen sobrecarga de presión y volumen, pérdida de músculo, enfermedad muscular primaria o demandas periféricas excesivas, como falla de alto rendimiento. En la forma habitual de insuficiencia cardíaca, el músculo cardíaco ha reducido la contractilidad. Esto produce una reducción en el gasto cardíaco, que luego se vuelve inadecuado para satisfacer las demandas periféricas del cuerpo. Los 4 determinantes principales del rendimiento del ventrículo izquierdo (VI) generalmente se alteran de la siguiente manera:

1. Hay una disminución intrínseca de la contractilidad muscular.
2. La precarga o la presión de llenado auricular izquierdo aumenta, lo que resulta en congestión pulmonar y disnea.
3. Aunque la presión arterial sistémica a menudo se reduce, hay un aumento en la resistencia vascular sistémica (poscarga), que puede reducir aún más el gasto cardíaco.
4. La frecuencia cardíaca generalmente aumenta como parte de un mecanismo compensatorio asociado con un aumento en el tono simpático y las catecolaminas circulantes.

En pacientes con enfermedad coronaria, a menudo hay un desequilibrio entre el suministro y la demanda de oxígeno del miocardio. Un aumento en el tamaño del corazón puede ser particularmente perjudicial al aumentar la tensión de la pared debido a la relación de Laplace y al aumento del consumo de oxígeno del miocardio[31].

Variabilidad del ritmo cardíaco (HRV)

Existe en el ámbito de la investigación un interés generalizado en el estudio de la variación de la temporalidad latido a latido del corazón, conocido como variabilidad de la frecuencia cardíaca (HRV por su nombre en inglés “Heart Rate Variability”). La fluctuación de la frecuencia cardíaca está relacionada con diversos trastornos cardiovasculares y se ha convertido en una herramienta ampliamente utilizada en la evaluación del comportamiento regular de la frecuencia cardíaca [1].

Definición de la HRV

La frecuencia cardíaca de referencia de un sujeto es generalmente como la frecuencia cardíaca (pico a pico) promediada durante 60 segundos HR_{60} , y esta varía entre pacientes dependiendo de la actividad y de su aptitud cardiovascular. Sin embargo, el intervalo RR instantáneo o la frecuencia cardíaca latido a latido (HR) también varía alrededor de la media de (HR_{60}).

Teniendo en cuenta que el promedio la frecuencia cardíaca es de aproximadamente 72 bpm, aunque parece haber una tendencia de esta media a posarse alrededor de 80 bpm y luego hasta aproximadamente 70 bpm en este segmento de 60 segundos. Este fenómeno de aumento y disminución de la frecuencia cardíaca en diferentes escalas de tiempo se conoce como variabilidad del ritmo cardíaco (HRV) también referido en la literatura como señal de interlatido cardíaco (NN).

Reflejos de control en la HRV

Los centros vago y simpático son estimulados simultáneamente por muchos receptores diferentes en los sistemas cardiovascular y nervioso central. Estos receptores, en general, tienen un efecto taquicárdico o bradicárdico en la frecuencia cardíaca.

Es importante notar que, aunque la frecuencia cardíaca es probablemente el mejor índice de equilibrio simpático-vagal, la pérdida de correlación entre la HRV y otros parámetros fisiológicos pueden tomarse como un signo de deterioro en el estado del paciente (como en el estado hemodinámico donde los correlatos son HR y HRV). Además, debido a que HRV tiende a ser función de la HR basal, podría ser ventajoso normalizar la HRV usando el HR para su análisis.

Técnicas de análisis lineal

Momentos Estadísticos

En matemáticas, un momento es una medida cuantitativa específica de la forma de una función. Si la función es una distribución de probabilidad, entonces el momento cero es la probabilidad total (es decir, uno), el primer momento es el valor esperado, el segundo momento central es la varianza, el tercer momento estandarizado es la asimetría y el cuarto momento estandarizado es la curtosis.

- El concepto matemático está estrechamente relacionado con el concepto de momento en física.
- En estadística, los valores sobre los cuales se trabaja ya no son masas como en su aplicación en la mecánica, pero aún miden características relativas a dato central relativas respecto al resto de los valores.

Los momentos estadísticos son parámetros que pueden ser utilizados para describir las características de la evolución temporal en fenómenos como, por ejemplo, los ciclos de la concentración plasmática y de la tasa de excreción urinaria que sigue a la administración de una dosis única de un medicamento [32]. Por esta razón, es una práctica común en el análisis lineal de series temporales realizar un ventaneo dinámico a la señal y obtener para cada ventana los valores de los momentos estadísticos para lograr observar si la evolución temporal de los datos obedece a un comportamiento esperado dentro de un régimen Gaussiano o se desvía del mismo.

Media

El primer momento hace referencia a la media aritmética de una distribución, esta variable representa el valor promedio de todas las observaciones en una muestra. Si las observaciones de la muestra de tamaño n , son x_1, x_2, \dots, x_n . Entonces la media de la muestra puede verse representada por la expresión:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{x_1 + x_2 + \dots + x_n}{n}. \quad (5.1)$$

Varianza

Una medida de tendencia central que captura la dispersión de los valores respecto a una valor medio es la varianza. Para medir este grado de dispersión, se utiliza la varianza, relacionada al segundo momento estadístico. La varianza se define como:

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (5.2)$$

Coeficiente de Variación

El coeficiente de variación, también denominado como coeficiente de variación de Pearson, es una medida estadística que brinda información acerca de la dispersión relativa de un conjunto de datos. La expresión para este coeficiente se muestra en la ecuación 5.3

$$CV = \frac{\sigma_i}{\bar{x}} \quad (5.3)$$

El coeficiente de variación se utiliza para comparar conjuntos de datos pertenecientes a poblaciones distintas, este tiene en cuenta el valor de la media. Por lo tanto, el coeficiente de variación permite contemplar las posibles distorsiones de las medias de dos o más poblaciones.

Skewness o Sesgo

El sesgo es un valor de error que mide la diferencia entre la esperanza matemática (media) y el valor real de la variable. También se conoce como el grado de asimetría de una distribución, si una curva de frecuencias se encuentra alineada a la izquierda se dice que tiene un sesgo positivo, mientras si esta alineamiento se dirige hacia la derecha, por el contrario, se dice que cuenta con un sesgo negativo. Este momento estadístico se define como muestra la ecuación 5.4.

$$s = \frac{1}{n} \frac{\sum(X - \bar{x})^3}{\sigma^3} \quad (5.4)$$

Así mismo la ecuación 5.5 representa el ajuste para el tercer momento muestral:

$$s = \frac{n}{(n-1)(n-2)} \frac{\sum(x - \bar{x})^3}{s^3} \quad (5.5)$$

Curtosis

La curtosis mide que tipo de gráfico tiene una distribución, comparando con una distribución de tipo normal, es conocida también como el grado de apuntamiento de una curva. Lo que significa que medirá Este momento se define en la ecuación 5.6.

$$k = \frac{1}{n} \frac{\sum(X - \bar{x})^4}{\sigma^4} \quad (5.6)$$

El ajuste muestral de la curtosis puede apreciarse en la ecuación 5.7:

$$k = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \frac{\sum(x - \bar{x})^4}{s^4} - \frac{3(n-1)^2}{(n-2)(n-3)} \quad (5.7)$$

Señal y Serie de Tiempo

Una señal se representa matemáticamente según una variable o más variables independientes. Una señal de audio, por ejemplo, se puede representar según la amplitud (intensidad) del valor en tensión captada por el dispositivo transductor y cómo evoluciona a lo largo del eje temporal. Las señales pueden describir una amplia variedad de fenómenos físicos. Las señales se representan matemáticamente como funciones de una o más variables independientes [33]. Las señales se pueden clasificar en analógicas, discretas en el tiempo y digitales. La diferencia principal entre estas señales es si son continuas o discretas en amplitud y tiempo. En la naturaleza, las señales son analógicas, pero para tratarlas computacionalmente pueden discretizarse en señales digitales.

Dado que una serie de tiempo es un conjunto de observaciones x_i , cada una de las cuales se registra en un tiempo específico t [34]. Generalmente una señal contiene dicho registro de temporalidad y por lo tanto puede denominarse una serie de tiempo.

Ventanas y Técnica de Ventaneo

En el contexto estricto de la matemática, las ventanas son funciones utilizadas para evitar las discontinuidades al principio y al final de los bloques analizados ó para realizar modificaciones o transformaciones a la información contenida dentro de dicha ventana.

En el procesamiento de señales, una ventana se utiliza cuando el análisis se centra en una señal o serie de tiempo de longitud voluntariamente limitada a un marco de observación. Esto se logra tomando una fracción de la señal original y aplicando las funciones ventana correspondientes.

El deslizamiento de esta ventana de observación a lo largo de la señal es a lo que se denomina **ventaneo** de una señal. Tomando en cuenta que existirá un **overlapping** ó superposición entre ventanas (una fracción de la señal que pertenece a la ventana anterior y una fracción de la ventana consecutiva) se considera que la señal conserva la continuidad de observación y aporta una mirada significativa a la evolución temporal de la función aplicada a lo largo de la señal. Un ejemplo de este procedimiento puede observarse en la figura 5.7, donde puede notarse el procedimiento anteriormente descrito sobre una señal a) de la cual se extrae una porción o ventana b) y se aplica una función o transformación a dicha ventana c). Este procedimiento continuará hasta alcanzar el final de la señal discretizada.

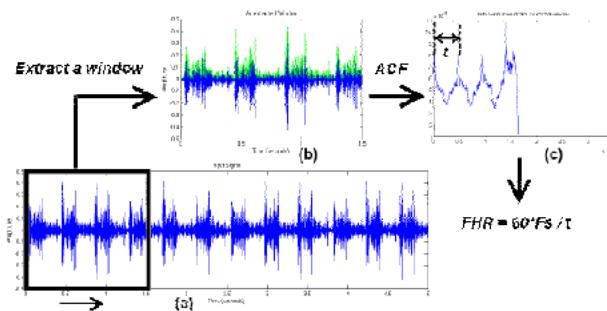


Figura 5.7: Ventaneo de una Señal

Técnicas de análisis No-Lineal

Sistema Complejo/s y Complejidad:

En la actualidad se ha encontrado cada vez más y más que existen ciertos comportamientos y características en común (Emergencia de auto-organización, multiescalamiento, autosimilitud) dentro de compendios de datos obtenidos de diversos sistemas multicomponentes, éste comportamiento que muestra señales de ser “universal” es generalmente denominado Complejidad, mientras que los sistemas que muestran éstas características son llamados Sistemas Complejos. A pesar de que aún hay muchas opiniones respecto a lo que es y comprende un sistema complejo, existen aspectos comunes que pueden ser reconocidos como signos de complejidad:

- Un sistema complejo tiene muchos grados de libertad; individuos, partículas, unidades. Cada uno de sus componentes es un nodo en una red con una compleja tipología de enlaces internodales, representando las interacciones no lineales entre cada unidad.
- La complejidad está asociada a la emergencia de comportamiento auto-organizado, es decir, en la dinámica no-lineal existe un mecanismo cooperativo que detona la formación espontánea de estructuras auto-organizadas.
- Los estados auto-organizados tienen características y escalas espacio-temporales difícilmente obtenidas de una función simple (lineal) respecto a intervención externa o procedimientos lineales de análisis de conjuntos.

Un sistema biológico es en consecuencia un sistema complejo tomando en cuenta todas las unidades que lo componen, sus grados de libertad y la emergencia de auto-organización.

Grupos de investigadores concentran sus esfuerzos en llegar a modelos de redes complejas que no sólo enlacen las funciones de órganos y tejidos, sino también llenar el espacio entre biología molecular y la fisiología del cuerpo humano aplicando conocimientos de campos como: Teoría de Redes, Física Estadística, Minería de Datos, Ciencias de la Información y Procesamiento de Señales [28].

Información y entropía de la información

Se entiende por información a toda aquello que conforma un documento de texto, grabación de audio o archivo de datos cuya medida en bits es el número de unos y ceros necesarios para almacenar el texto, el sonido o los datos utilizando la codificación digital más eficiente posible. Esta definición está formalizada por la teoría algorítmica de la información que define a la información algorítmica como una secuencia arbitraria s de ceros y unos que es posible contener estructurar por medio de un algoritmo $K(s)$ que posee el tamaño, en bits, mínimo ejecutable para imprimir la secuencia.

Para conocer si existen métodos en los que la información que forma una serie de tiempo se ordena siguiendo comportamientos particulares, es necesario introducir medidas que pertenecen al paradigma de la Teoría de la Información.

Desde el punto de vista de la termodinámica, la entropía es una cantidad asociada a un sistema en estado de equilibrio. Así la entropía de un sistema puede cambiar al intercambiar calor o trabajo con sus alrededores. Para procesos irreversibles, la entropía debe obedecer la segunda ley que dice que la entropía de un sistema aislado nunca puede disminuir. La definición convencional de entropía en mecánica estadística se debe a Boltzmann y Gibbs, donde el tratamiento de cantidades extensivas como el número de partículas o la energía con valores definidos para microestados individuales (en mecánica cuántica, un microestado se especifica dando una lista completa de los números cuánticos del sistema y en la mecánica clásica, es la ubicación en el espacio de fase - las posiciones y momentos de todas las partículas del sistema). Pero, la entropía de Gibbs se define sólo para conjuntos estadísticos o ensambles de partículas [35].

El concepto de entropía se ha aceptado generalmente como una medida de variabilidad o “desorden” de un sistema. No obstante, para cuantificarla es necesario distinguir que si los conjuntos de datos que conforman a los sistemas de interés fueran infinitos y perfectos, sería posible determinar un valor preciso de entropía, sin embargo para conjuntos de datos biológicos estas condiciones difícilmente se cumplen, dado que son sistemas dinámicos. Para ello, la teoría plantea que incluso una estimación imperfecta ó aproximada de la entropía podría usarse para clasificar conjuntos de series de tiempo en su jerarquía de orden siguiendo la idea de que las series de tiempo con elementos repetidos surgen de sistemas más ordenados y se caracterizarían razonablemente por un bajo valor de entropía [36].

Entropía Aproximada

La medida de entropía aproximada surge debido al planteamiento anteriormente mencionado de obtener una medida de entropía que se ajuste a la tasa media de producción de información de un sistema.

La entropía aproximada se expresa como el logaritmo natural negativo de la probabilidad condicional de que una ventana corta de los datos, o “plantilla”, se repita durante la serie de tiempo. Ya habiendo seleccionado una plantilla de tamaño m , se identifica otras plantillas que son arbitrariamente similares y se determina cuáles de ellas permanecen arbitrariamente similares para el siguiente punto $m + m_0$. “Arbitrariamente similar” significa que los puntos están dentro de un margen de tolerancia r entre sí, donde r normalmente se selecciona como un factor de la desviación estándar. El logaritmo negativo de la probabilidad condicional se calcula para cada plantilla posible y se promedian los resultados.

Si los datos están ordenados, entonces las plantillas que tienen formas similares para m puntos son a menudo de forma similares para $m + 1$ puntos, la probabilidad condicional se aproxima a 1 y el logaritmo negativo y la entropía se acercan a 0 (alta regularidad).

Este cálculo de la Entropía Aproximada se expresa mediante la ecuación 5.8:

$$AppEn \approx -\sum log(1 + A_i)/(1 + B_i) \quad (5.8)$$

Entropía Muestral

A pesar de la utilidad potencial de utilizar una medida de entropía aproximada existen cuestiones prácticas de gran importancia en su implementación. Una de ellas se presenta al preguntar ¿Qué sucede

si algunas plantillas no tienen coincidencias y las probabilidades condicionales no están definidas?

Uno de los acercamientos a esta cuestión da como resultado a la Entropía Muestral, ya que permite que las plantillas coincidan entre sí. De modo que, si no existen coincidencias, la probabilidad condicional es 1 y la medida de la *AppEn* es 0, esto sería un indicador de un orden “perfecto”. Si es el caso que existan algunas coincidencias entre plantillas, entonces el resultado está sesgado hacia 0, y el sesgo se resuelve con alargamiento de conjuntos de datos y más coincidencias de plantillas. Por tanto, si la serie de tiempo A surge de un sistema más ordenado que la serie de tiempo B, entonces *AppEn* de A siempre será menor que *AppEn* de B independientemente del número de coincidencias de plantilla y la extensión del sesgo [37].

El algoritmo para la entropía muestral consiste simplemente en contar los números A y B, lo que requiere verificar la distancia entre los $[N(N - 1)]/2$ pares de puntos, luego contar el de vectores que coinciden con m y $m + 1$ puntos. De manera similar a la entropía aproximada, el valor de la entropía muestral estará dado por el logaritmo negativo como lo muestra la ecuación 5.9:

$$SampEn = -\log((\Sigma A_i / \Sigma B_i)) = -\log(A/B) \quad (5.9)$$

Entropía Espectral

La entropía espectral de una señal es una medida de su distribución de potencia espectral. El concepto se basa en la entropía de la información o entropía de Shannon, trata al distribución de energía normalizada de la señal en el dominio de la frecuencia como una distribución de probabilidad, y calcula la entropía de Shannon, que en este contexto es la entropía espectral de la señal. Esta propiedad suele ser útil para la extracción de características en la detección y diagnóstico de irregularidades o errores por lo que es ampliamente utilizado como una característica en el reconocimiento de voz y procesamiento de señales biomédicas.

Las ecuaciones para la entropía espectral surgen de las ecuaciones para el espectro de potencia y la distribución de probabilidad para una señal.

Para una señal $x(n)$ su espectro de potencia es $S(m) = |X(m)|^2$ donde $X(m)$ es la transformación discreta de Fourier de $x(n)$, al final el cálculo de la entropía espectral al tiempo t es:

$$H(t) = -\sum_{m=1}^N P(t, m) \log_2 P(t, m) \quad (5.10)$$

Objetos fractales y procesos auto-similares

El concepto de fractal se asocia más a menudo con objetos geométricos que satisfacen dos criterios: auto-semejanza y dimensionalidad fraccionaria. La autosimilitud significa que un objeto se compone de subunidades y subunidades en múltiples niveles que (estadísticamente) se asemejan a la estructura del objeto completo [38].

Matemáticamente, esta propiedad debería mantenerse en todas las escalas. Sin embargo, en el mundo real, necesariamente existen límites inferiores y superiores sobre los que se aplica tal comportamiento auto-similar. El segundo criterio para un objeto fractal es que tenga una dimensión fraccionaria. Este requisito distingue a los fractales de los objetos euclidianos, que tienen dimensiones enteras.

Es posible generar estos objetos dadas sus propiedades matemáticas, ejemplo de ello se muestra en la figura 5.8 que muestra un fractal conocido generado por medio de la iteración de la expresión:

$$X1 := 0,85 * X0 + 0,04 * Y0; Y1 := -0,04 * X0 + 0,85 * Y0 + 1,60;$$

A este fractal se le conoce como el Helecho de Barnsley en honor al matemático británico Michael Barnsley.



Figura 5.8: Helecho de Barnsley: Objeto fractal generado por computadora

El concepto de estructura fractal, que carece de una escala de longitud característica, puede extenderse al análisis de procesos temporales complejos. Sin embargo, un desafío en la detección y cuantificación de escalas auto-similares en series de tiempo complejas es el siguiente: aunque las series de tiempo generalmente se trazan en una superficie bidimensional, una serie de tiempo en realidad involucra dos variables físicas diferentes[39].

Un proceso dependiente del tiempo (o serie de tiempo) es auto-similar si de acuerdo con la ecuación 5.11, las propiedades estadísticas de ambos lados de la ecuación son idénticas. En otras palabras, un proceso auto-similar, $y(t)$, con un parámetro α (llamado el parámetro de auto-similitud) tiene la misma distribución de probabilidad que un proceso correctamente re-escalado.

$$y(t) \equiv a^\alpha y\left(\frac{t}{a}\right) \quad (5.11)$$

La figura 5.9 muestra el concepto de auto-similitud para un recorrido aleatorio simulado de una serie de tiempo. (a) Se muestran dos ventanas de observación, con escalas de tiempo n_1 y n_2 , para una serie de tiempo auto-similar $y(t)$. (b) La ampliación de la ventana más pequeña con escala de tiempo n_1 denota que las fluctuaciones en (a) y (b) parecen similares siempre que dos factores de ampliación diferentes, M_x y M_y , se aplican en las escalas horizontal y vertical, respectivamente. (c) La distribución de probabilidad, $P(y)$, de la variable y para las dos ventanas en (a), donde s_1 y s_2 indican las desviaciones estándar para estas dos funciones de distribución. (d) Gráfico $\log - \log$ de las escalas características de fluctuaciones s versus los tamaños de ventana n .

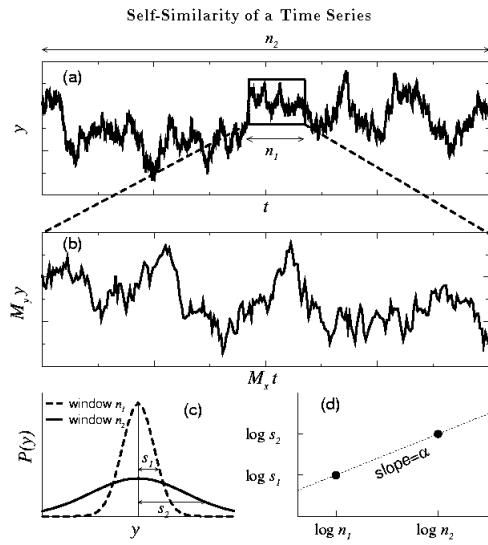


Figura 5.9: Auto-similitud de una Serie de Tiempo

El parámetro de auto-similitud α definido en la ecuación 5.11 se puede calcular mediante una relación simple dada por la ecuación 5.12 donde M_x y M_y son los factores de aumento apropiados a lo largo de la dirección horizontal y vertical, respectivamente.

$$\alpha = \frac{\ln M_y}{\ln M_x} \quad (5.12)$$

Esta relación es simplemente la pendiente de la línea que une los puntos (n_1, s_1) y (n_2, s_2) en una gráfica logarítmica .

Dimensión Fractal de Higuchi

La dimensión de Higuchi es un valor aproximado para la dimensión de conteo de cajas aplicado a una serie de tiempo, el valor es obtenido en pocas palabras via la aproximación logarítmica por lo que tambien se entiende como “Método Higuchi”. Dada la serie de tiempo X consistente de N datos y un parámetro k mayor o igual a 2 se calcula entonces a partir de la ecuación:

$$L_m(k) = \frac{N-1}{[\frac{N-m}{k}]k^2} \sum_{i=1}^{[\frac{N-m}{k}]} |X_N(m+ik) - X_N(m+(i-1)k)| \quad (5.13)$$

Donde la longitud L_k es definida por el promedio de las longitudes k ; $L_1(k), \dots, L_k(k)$

$$L(k) = \frac{1}{k} \sum_{m=1}^k L_m(k) \quad (5.14)$$

Y finalmente el valor de la dimensión fractal de higuchi de la serie de tiempo X es generado a partir de la pendiente de la función lineal mejor adaptada a los puntos pleteados sobre $\{(\log \frac{1}{k}, \log L(k))\}$

Exponente de Hurst

El exponente de Hurst es usado como una medida de dependencia a largo plazo en una serie de tiempo, lo cual se relaciona con la auto-correlación de la serie y el ritmo al cual ésta se reduce mientras que se separan más pares de valores. Tiene una relación directa con la dimensión fractal, midiendo la mesurada o alta aleatoriedad de una serie de datos.

Comprende valores entre 0 y 1, estableciendo rangos en los que describe de diferentes maneras a la serie de tiempo; valores de 0.5 indican que el comportamiento se refiere a un movimiento Browniano, definido como el movimiento aleatorio de partículas suspendidas en un medio, lo que nos indicaría que no existe una tendencia real hacia la cual la serie de tiempo se incline con los datos proporcionados y se espera un comportamiento aleatorio. Un valor mayor a 0.5 indica una inclinación positiva hacia la tendencia que tienen los datos, esperando a diferentes grados mantenerse en un comportamiento futuro similar al que se dió en el pasado; finalmente valores menores a 0.5 indican una correlación negativa a la tendencia, con lo que en grados de magnitud se podría esperar un comportamiento opuesto al previamente observado en el conjunto de datos evaluado.

Análisis fractal y series de tiempo con procesos auto-similares

Para un proceso auto-similar con $\alpha > 0$, las fluctuaciones crecen con el tamaño de la ventana en forma de ley de potencia. Por lo tanto, las fluctuaciones en las ventanas de observación grandes son exponencialmente mayores que las de las ventanas más pequeñas. Como resultado, la serie de tiempo no tiene límites.

Sin embargo, para la mayoría de las series de tiempo fisiológicas de interés, como la frecuencia cardíaca, están limitadas; no pueden tener amplitudes arbitrariamente grandes, independientemente de la longitud del conjunto de datos. Esta restricción práctica causa más complicaciones para los análisis considerando un acercamiento a un subconjunto de la serie de tiempo, un patrón aparentemente auto-similar.

Ruido Rosa ó Ruido 1/f

El ruido rosa o ruido $1/f$ es una señal o proceso con un espectro de frecuencia tal que la densidad espectral de potencia (potencia por intervalo de frecuencia) es inversamente proporcional a la frecuencia de la señal de forma:

$$S(f) \propto \frac{1}{f^\alpha}$$

Las fluctuaciones de $1/f$ se encuentran ampliamente en la naturaleza. A partir de procesos de memoria larga con correlaciones a largo plazo y $1/f^\alpha$ (con $0.5 \lesssim \alpha \lesssim 1.5$) comportamiento de espectros de potencia a bajas frecuencias f se ha observado en física, tecnología, biología, astrofísica, geofísica, economía, psicología, lenguaje e incluso música.

Análisis de fluctuación sin tendencia (DFA)

Una serie de tiempo limitada se puede mapear a un proceso auto-similar por integración. Sin embargo, otro desafío al aplicar este tipo de análisis fractal a los datos fisiológicos es que estas series de tiempo son a menudo muy no estacionarias. Una definición general y simplificada caracteriza a una serie de tiempo como estacionaria si la media, la desviación estándar y los momentos superiores, así como las funciones de correlación son invariantes bajo la traducción del tiempo. Las señales que no obedecen estas condiciones no son estacionarias y el procedimiento de integración solo exagerará aún más la no estacionariedad de los datos originales [40].

Para superar esta complicación, se modificó un método derivado del análisis de la raíz cuadrada media, llamado análisis de fluctuación sin tendencia DFA, el cuál permite la exploración de correlaciones a corto y largo plazo en series de variabilidad de frecuencia cardíaca no estacionarias[13].

El DFA cuantifica las propiedades de autocorrelación fractal de las señales. Es útil para analizar señales biomédicas que en su mayoría son complejas y no estacionarias. En este artículo revisamos ejemplos seleccionados de aplicación del método DFA en cardiología, neurología y otros estudios. También presentamos nuestros hallazgos, algunos de nuestros trabajos originales. Concluimos que usando el método DFA podemos determinar qué señal es más regular y menos compleja (en la práctica para distinguir sujetos sanos de enfermos)[41].

Para aplicar el algoritmo DFA en una señal unidimensional $B(i)$, $i = 1, \dots, N$. Primero, la señal integrada se calcula de acuerdo con la ecuación 5.15:

$$y(k) = \sum_{i=1}^k (B(i) - B_{\text{avg}}) \quad (5.15)$$

donde B_{avg} es el valor medio de la señal. A continuación, los datos se dividen en segmentos de longitud n y se encuentra la aproximación lineal y_n utilizando un ajuste por mínimos cuadrados en cada segmento por separado (lo que representa la tendencia en una sección determinada).

La fluctuación promedio $F(n)$ de la señal alrededor de la tendencia viene dada por la ecuación 5.16:

$$F(n) = \sqrt{\frac{1}{N} \sum_{k=1}^N (y(k) - y_n(k))^2} \quad (5.16)$$

Los cálculos se repiten para todos los n considerados. Es útil notar que, en general, $F(n)$ aumentará con el tamaño del segmento n . Finalmente, se crea un gráfico logarítmico doble $\log F(n)$ vs $\log n$ siendo el relación entre $F(n)$ y el tamaño del segmento n donde reside el interés. La dependencia lineal indica la presencia de fluctuaciones propias y la pendiente de la línea $F(n)$ determina el exponente de escala α que se aproximan por una ley de potencias como muestra la ecuación 5.17:

$$F(n) \sim n^\alpha \quad (5.17)$$

Las ventajas de DFA sobre los métodos convencionales (por ejemplo, análisis espectral y análisis de Hurst) son que permite la detección de auto-similitud intrínseca incrustada en una serie de tiempo aparentemente no estacionaria, y también evita la detección espuria de auto-similitud aparente, que puede ser un problema.

Análisis por Gráficos de Poincarè

El gráfico de Poincaré, también llamado mapa de retorno, es la técnica más simple para describir la dinámica no lineal de un fenómeno. Esta técnica visual para reconocer los patrones de correlación ocultos de una serie de tiempo explota las propiedades del plano de fase para sistemas bidimensionales y proporciona un método conveniente para descubrir la existencia de comportamientos periódicos o posiblemente más complejos, como el movimiento cuasiperiódico o caótico.

La técnica popular para medir la gráfica de Poincaré es ajustar una elipse y medir la dispersión de puntos a lo largo del eje menor (SD1) y el eje mayor (SD2). Estos descriptores estándar del gráfico de Poincaré son vectores que se utilizan para cuantificar la gráfica que mide la variabilidad bruta de los datos de la serie temporal al comparar el comportamiento de ciertos datos distribuidos en el plano-fase. La figura 5.10 muestra la implementación de un Gráfico de Poincarè sobre datos generados aleatoriamente con sus respectivos descriptores.

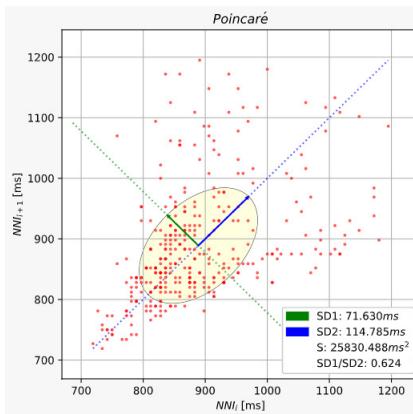


Figura 5.10: Gráfico de Poincarè de datos aleatorios

En el análisis de HRV, el descriptor estándar SD1 se considera que describe la variabilidad a corto plazo y SD2 la variabilidad a largo plazo sin embargo, la falta de información temporal es la principal limitación de los descriptores estándar del gráfico de Poincaré. Para el análisis de la señal a largo plazo, la inspección de la variación temporal de cualquier medida da más información que la medida global. Por lo tanto, animar la progresión temporal de SD1 y SD2 para diferentes condiciones fisiológicas puede guiar la interpretación de los métodos cuantitativos y dar una idea del grado de dinámica temporal [42].

Debido a la alta dimensionalidad de la señal cardiaca la búsqueda de un atractor (como suele hacerse en espacios-fase) se ve limitada, razón por la cual se considera útil hacer el mapa de retorno para la misma ventana de tiempo definida para el resto de las técnicas, y graficar la evolución del cociente $R = \frac{SD1}{SD2}$ esto para verificar la existencia de fenómenos persistentes dentro la señal.

Procesamiento de señales con wavelets

Wavelets

Las wavelets u ondeletas son funciones ampliamente utilizadas en el procesamiento digital de señales para descomponer y modelar señales. De manera análoga a como la transformada de Fourier descompone una señal en una familia de sinusoides complejos, la transformada de ondeletas descompone una señal en una familia de ondeletas. A diferencia de las sinusoides, que son simétricas, suaves y regulares, las ondeletas pueden ser de morfologías variadas, ya sean simétricas o asimétricas, nítidas o suaves, regulares o irregulares.

Se han definido un amplio número de ondeletas cuyas formas varían de acuerdo a la aplicación. Por ejemplo la aplicación del análisis con ondeletas de Morlet se utiliza para discriminar el comportamiento anormal de los latidos del corazón en el ECG [43]. Esta ondeleta está definida matemáticamente como:

$$\Psi_\sigma(t) = c_\sigma \pi^{-\frac{1}{4}} e^{-\frac{1}{2}t^2} \left(e^{i\sigma t} - \kappa_\sigma \right)$$

Al ver la forma de dicha ondeleta es clara la razón de su uso en este problema debido a su semejanza con un complejo QRS como puede verse en la figura 5.11.

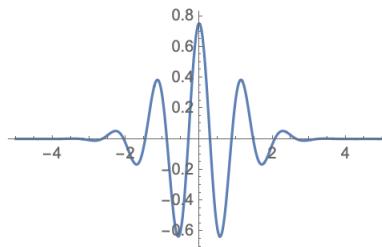


Figura 5.11: Ondeleta de Morlet

La Transformada Wavelet

La transformada wavelet (WT) ó transformada de ondeletas calcula los productos internos de una señal mediante una familia de ondeletas [44]. Las herramientas de transformación de se clasifican en herramientas de ondeletas continuas y herramientas de ondeletas discretas. Por lo general, utiliza las herramientas de ondeletas continuas para el análisis de señales, como el análisis de auto-similitud y el análisis de tiempo-frecuencia.

La transformada de ondeleta continua (CWT) está definida en términos de dilataciones y traslaciones de un prototipo o función madre $\Phi(t)$, en el dominio del tiempo y el dominio de Fourier la transformada está representada por:

$$\psi_{ab}(t) \frac{1}{\sqrt{a}} \psi \left(\frac{t-b}{a} \right) \leftrightarrow \psi_{ab}(\Omega) = \sqrt{a} \psi(a\Omega) e^{-jb\Omega}$$

La transformada de ondeletas discretas (DWT) es el resultado de la discretización directa de la CWT y es ampliamente utilizada para el análisis y el procesamiento de señales, como la reducción de ruido, la compresión de datos, la detección de picos, etc. Las propiedades de las ondeletas y la flexibilidad para seleccionar ondeletas hacen que sean una herramienta beneficiosa para aplicaciones de extracción de características.

Densidad espectral de potencia de ondeletas

La densidad espectral de potencia de la ondeleta es un equivalente de baja resolución a la densidad espectral de potencia tradicional basada en la transformada de Fourier. La información de tiempo obtenida por la WT se utiliza para eliminar los componentes de alta frecuencia de la señal analítica que interferirían con el análisis del ruido de línea base.

Los coeficientes de la WT proporcionan información sobre la correlación entre la ondeleta (a una determinada escala) y la matriz de datos (en una ubicación particular). Una amplitud positiva mayor implica una correlación positiva mayor, mientras que una amplitud negativa grande implica una correlación negativa alta [45].

Una forma útil de determinar la distribución de energía dentro de la matriz de datos es trazar la potencia de la ondeleta, equivalente a la amplitud al cuadrado. Al buscar regiones dentro del espectro de potencia Wavelet (WPS) una región de gran potencia puede determinar qué características de su señal son importantes y cuáles pueden ignorarse. Ejemplo de esto es la figura 5.12, la cual muestra una aplicación de este método en la determinación de comportamientos fractales en los mercados financieros [46].

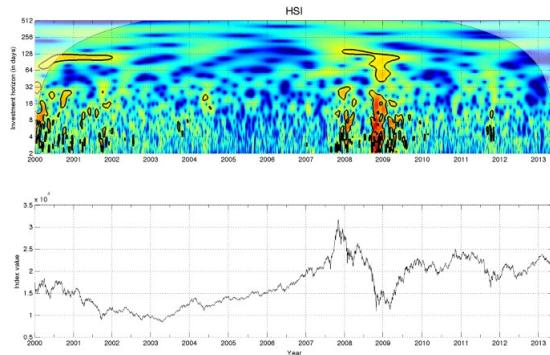


Figura 5.12: WPS de serie de tiempo relacionada con el índice HSI

Clasificación por medio de Machine Learning

El paradigma de Machine Learning

Se conoce como Machine Learning (o Aprendizaje Automático en el idioma español) al paradigma computacional que incluye un conjunto de métodos que permiten a las computadoras “aprender” de los datos para hacer y mejorar predicciones e inferencias.

El ML representa un cambio de paradigma en la programación convencional y da paso a la programación “indirecta”. Para la programación convencional la resolución de un problema dado consiste en el diseño de programas que alimentan instrucciones explícitamente a la computadora para tratar los datos. La particularidad de la programación indirecta es que la resolución del problema se lleva a cabo mediante la provisión de datos a la computadora y los modelos que dan respuesta son generados a partir de métodos que cumplen la función de realizar inferencias con los datos, dando la sensación de que las máquinas aprenden por sí mismas sin ser programadas.

Para comprender en su totalidad este paradigma de la programación es necesario definir los siguientes términos[47]:

Algoritmo

En términos generales a un método preciso adecuado al dispositivo que se utiliza para resolver un problema determinado en un número finito de pasos ,que frecuentemente implica la repetición de una operación, se conoce como algoritmo[48],[21]. Un algoritmo es un conjunto de reglas y procedimientos matemáticos que sigue una máquina para lograr un objetivo en particular. La analogía más explicativa de un algoritmo es la que lo plantea como una receta que define las entradas, la salida y todos los pasos necesarios para pasar de las entradas a la salida, en este caso de datos a información.

Algoritmos Supervisados y No Supervisados

Existe en ML una distinción importante entre la modalidad de los algoritmos: Supervisados y no Supervisados.

Los algoritmos supervisados “aprenden” de los datos introducidos por una persona, esto implica que es necesaria la intervención humana para etiquetar, clasificar e introducir los datos en el algoritmo de acuerdo a las clases en las que se desea clasificar los datos. Por ende, este genera datos de salida esperados, ya que a la entrada han sido etiquetados previamente. Por ello se dice que se conoce el objetivo a priori antes de realizar una inferencia sobre los datos.

Para los algoritmos no supervisados, a diferencia del caso anterior, no existe la intervención humana en los datos. Estos algoritmos “aprenden de datos” con elementos no etiquetados buscando patrones o relaciones entre ellos mediante una estructura mucho más compleja y haciendo uso de modelos matemáticos de mayor impacto. Dado que estos algoritmos suelen contener múltiples etapas retro alimentadas entre ellas, su uso suele conllevar un costo computacional elevado con respecto a su contra parte supervisada.

Modelo de Aprendizaje

Un modelo de aprendizaje ó algoritmo de ML es el programa utilizado para aprender, a partir de datos, una asignación de las entradas a las predicciones. Puede ser un conjunto de pesos para un modelo lineal o para una red neuronal. En su idioma original a estos modelos se les conoce como “learner”, que puede interpretarse como una unidad de aprendizaje.

Otros nombres para la palabra “modelo”, dependiendo de la tarea son clasificador o modelo de regresión. El modelo de aprendizaje automático entrenado se denota matemáticamente como \hat{f} ó $\hat{f}(x)$.

Dataframe

Un Dataframe o Dataset es una tabla con los datos de los que aprende la máquina. El conjunto de datos contiene las características y el objetivo a predecir. Cuando se usa para inducir el entrenamiento de un modelo, el conjunto de datos generalmente se divide en dos grupos proporcionalmente distintos a los que se denomina datos de entrenamiento ó Training Set y datos de prueba o Testing Set.

Instancia

Una instancia es una fila en el conjunto de datos. Otros nombres para instancia son: punto (de datos), ejemplo, observación. Una instancia consta de los valores de la característica $x^{(i)}$ y, si se conoce, el resultado objetivo y_i .

Característica

Las características son las entradas utilizadas para la predicción o clasificación. Una característica es una columna en el conjunto de datos. A lo largo del trabajo, se supone que las características son interpretables, lo que significa que es fácil comprender lo que significan, como la temperatura en un día determinado o la altura de una persona. La interpretabilidad de las características es una gran suposición. Pero si es difícil comprender las características de entrada, es aún más difícil comprender qué hace el modelo. La matriz con todas las características se llama X y $x^{(i)}$ para una sola instancia. El vector de una sola característica para todas las instancias es x_j

Objetivo

El objetivo ó “Target” es la información que la máquina aprende a predecir. En notación matemática, el objetivo generalmente se llama y ó y_i para una sola instancia.

Tarea

Una tarea de ML es la combinación de un conjunto de datos con características y un objetivo. Dependiendo del tipo de objetivo, la tarea puede ser, por ejemplo, clasificación, regresión, análisis de supervivencia, agrupación o detección de valores atípicos ó “outliers”.

Predicción

La Predicción es lo que el modelo de aprendizaje automático “infiere”, lo cuál debería ser el valor objetivo en función de las características dadas. La predicción del modelo se denota por $\hat{f}(x^{(i)})$ ó \hat{y} .

Clasificadores Binarios

Un modelo de clasificación binaria es aquel que realiza la tarea de clasificar los elementos de un conjunto en dos grupos sobre la base de una regla de clasificación . Los problemas típicos de clasificación binaria incluyen:

Pruebas médicas para determinar si un paciente tiene cierta enfermedad o no; Control de calidad en la industria, decidiendo si se ha cumplido una especificación; En la recuperación de información, decidir si una página debe estar en el conjunto de resultados de una búsqueda o no. La clasificación binaria es la dicotomización aplicada a una situación práctica. En muchos problemas prácticos de clasificación binaria, los dos grupos no son simétricos y, más que la precisión general, es de interés la proporción relativa de diferentes tipos de errores .

Por ejemplo, en las pruebas médicas, detectar una enfermedad cuando no está presente (un falso positivo) se considera diferente a no detectar una enfermedad cuando está presente (un falso negativo).

Evaluación de Clasificadores Binarios

Existen varias métricas que se pueden utilizar para cuantificar el desempeño de un clasificador.

En medicina, a menudo se utilizan la sensibilidad y la especificidad, mientras que en la recuperación de información se prefieren la precisión y la exhaustividad.

Estas métricas pueden subdividirse a su vez en dos grupos principales que poseen propiedades muy diferentes entre sí:

- Las que son independientes de la frecuencia con la que ocurre cada categoría de eventos en la población (conocida como prevalencia).
- Las métricas que dependen de la prevalencia.

Dada una clasificación de un conjunto de datos específico, hay cuatro combinaciones básicas de categoría de datos real y categoría asignada: verdaderos positivos TP (asignaciones positivas correctas), verdaderos negativos TN (asignaciones negativas correctas), falsos positivos FP (asignaciones positivas incorrectas), y falsos negativos FN (asignaciones negativas incorrectas).

Estos valores pueden organizarse por medio de una tabla de contingencia como se muestra en la figura 5.1, la cual se emplea generalmente para registrar y analizar la asociación entre dos o más variables.

	Condición Positiva	Condición Negativa
Resultado Positivo	Verdaderos Positivos (TP)	Falsos Positivos (FP)
Resultado Negativo	Verdaderos Negativos (TN)	Falsos Negativos (FN)

Cuadro 5.1: Tabla de Contingencia para clasificadores binarios

Matríg de Confusión

Una matríg de confusión es una adaptación de la tabla de contingencia que sirve como una ayuda visual para evaluar el desempeño de un algoritmo de ML. Para ello se vale los resultados obtenidos por el modelo respecto al Testing Set. Cada columna de la matriz representa el númer de predicciones de cada clase, mientras que cada fila representa a las instancias en la clase real. En términos prácticos, presenta de manera visual los tipos de aciertos y errores que resultaron del proceso de aprendizaje del mismo con los datos como demuestra la figura 5.13.

		VALORES PREDICIÓN	
		Verdaderos positivos	Falsos Positivos
VALORES REALES	Falsos Negativos		Verdaderos Negativos

Figura 5.13: Matríg de confusión simple

Es importante notar que, una matriz la información que proporciona la diagonal de esta viene del hecho de que es similar a una matriz diagonal, o sea que posee n vectores propios linealmente independientes, lo que podría ser de ayuda al tratar de determinar la movilidad de los datos de una clase a otra.

Exactitud

Exactitud o “Accuracy” se refiere a la cercanía del resultado de una medición del valor verdadero. En términos estadísticos, la exactitud está relacionada con el sesgo de una estimación. Se conoce alternativamente como Tasa Verdadero Positivo o “True positive rate” y se representa mediante la proporción entre los positivos reales predichos por el algoritmo y todos los casos positivos.

En forma práctica la Exactitud es la cantidad de predicciones positivas que fueron correctas determinadas por:

$$(VP + VN) / (VP + FP + FN + VN)$$

La Precisión

La Precisión es una cantidad que se refiere a la dispersión del conjunto de valores obtenidos a partir de mediciones repetidas de una magnitud; en este sentido, cuanto menor es la dispersión mayor será la precisión. Se representa por la proporción entre el númer de predicciones correctas (tanto positivas como negativas) y el total de predicciones. En forma práctica es el porcentaje de casos positivos detectados dada por:

$$VP / (VP + FP)$$

Sesgo

El sesgo, también llamado “Bias” o “Inaccuracy” en este contexto pertenece al concepto de exactitud y representa la diferencia entre el valor medio y el verdadero valor de la magnitud medida

Sensibilidad y Especificidad

La sensibilidad y la especificidad son dos valores que indican la capacidad de un modelo clasificador para discriminar entre los casos positivos y los negativos. La Sensibilidad ó recuperación (“Recall” o “Sensitivity”), también se conoce como Tasa de Verdaderos Positivos (True Positive Rate) ó TP. Es la proporción de casos positivos que fueron correctamente identificadas por el algoritmo dada por:

$$VP/(VP + FN)$$

La Especificidad ó “Especificity” también conocida como la Tasa de Verdaderos Negativos, (“true negative rate”) o TN. Se trata de los casos negativos que el algoritmo ha clasificado correctamente. Expresa en términos cuantitativos la capacidad del modelo para detectar una clase en específico y está dada por:

$$VN/(VN + FP)$$

Curvas ROC y curvas de recuperación-precisión

Una herramienta útil para predecir la probabilidad de un resultado binario es la curva de Características Operativas del receptor, o curva ROC.

Este es un gráfico de la tasa de falsos positivos (eje x) frente a la tasa de verdaderos positivos (eje y) para varios valores de umbral candidatos diferentes entre 0,0 y 1,0 como muestra la figura 5.14, generalmente esta evaluación se hace a un solo modelo comparándolo contra un modelo sin habilidad es uno que no puede discriminar entre las clases y estimará una clase aleatoria o una clase constante en todos los casos. La línea sin habilidades cambia según la distribución de las clases positivas a negativas.

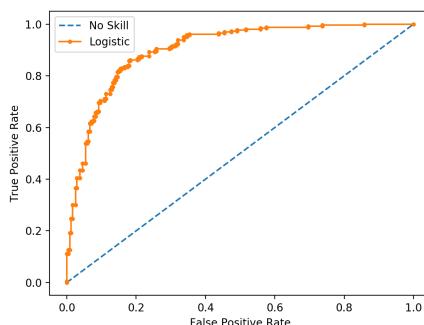


Figura 5.14: Curva ROC de un regresión logística

Otro enfoque similar mide la precisión y la recuperación para evaluar modelos de clasificación binaria.

La revisión tanto de la precisión como la memoria es útil en los casos en que existe un desequilibrio en las observaciones entre las dos clases. Específicamente, si existe un mayor número de ejemplos de ningún

evento (clase 0) y solo algunos ejemplos de un evento (clase 1). La razón de ello es que típicamente la gran cantidad de ejemplos de clase 0 significa que se tendría un menor interés en la habilidad del modelo para predecir correctamente la clase 0, produciendo un alto número de verdaderos negativos.

Es de notar que el cálculo de la precisión y la recuperación es que los cálculos no utilizan los verdaderos negativos sino únicamente se ocupa de la predicción correcta de la clase minoritaria, la clase 1.

Una curva de recuperación de precisión es un gráfico de la precisión (eje y) y la recuperación (eje x) para diferentes umbrales, de manera similar a la curva ROC como se muestra en la figura 5.15. Es una línea horizontal con el valor de la proporción de casos positivos en el conjunto de datos. Para un conjunto de datos equilibrado, esto es 0.5.

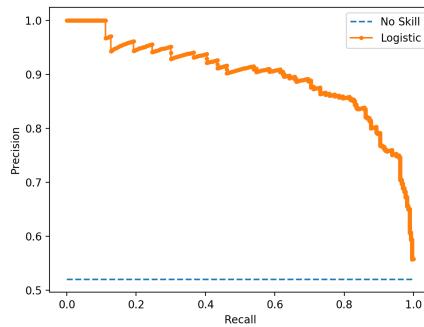


Figura 5.15: Curva RC de un regresión logística

Arboles de Decisión

En la toma de decisiones, se puede utilizar un árbol de decisiones para representar de forma visual y explícita las decisiones y la toma de decisiones. En el contexto de ML, los árboles de decisión son modelos que utilizan un sistema de toma de decisiones en forma de árbol para derivar una estrategia que pretende alcanzar una clasificación haciendo inferencias basadas en dichas decisiones respecto a los datos.

Extremely Randomized Trees Classifier

Es un tipo de algoritmo de aprendizaje supervisado el cual agrega los resultados de múltiples de-correlacionados arboles de decisión en una estructura de bosque que permite brindar clasificaciones a un grupo de valores de entrada.

Cada árbol de decisión contenido en el bosque es construido de la muestra original de entrenamiento, es entonces cuando en cada nodo de prueba, cada árbol adquiere una muestra randomizada de k características a partir de la cuál el árbol decide la mejor de todas para bifurcar los datos con el uso de un criterio matemático (típicamente el coeficiente de Gini). Este proceso da origen a los múltiples árboles de decisión de-correlacionados.

Para poder escoger entre la mejor característica, cada una es ordenada en forma descendente de acuerdo a el cálculo del coficiente de Gini para la muestra.

Análisis de Componentes Principales

El análisis de componentes principales (PCA) es una de las técnicas de reducción de dimensiones más utilizadas para transformar un conjunto de datos de gran tamaño en un conjunto de datos más pequeño, mediante la identificación de correlaciones y patrones con la preservación de la mayor parte de la información principal característica de los datos. Este método disminuye la complejidad del modelo y aumenta la eficiencia computacional. En la Figura 5.16 se muestra un ejemplo de reducción dimensional de 3 dimensiones a 3 componentes principales.

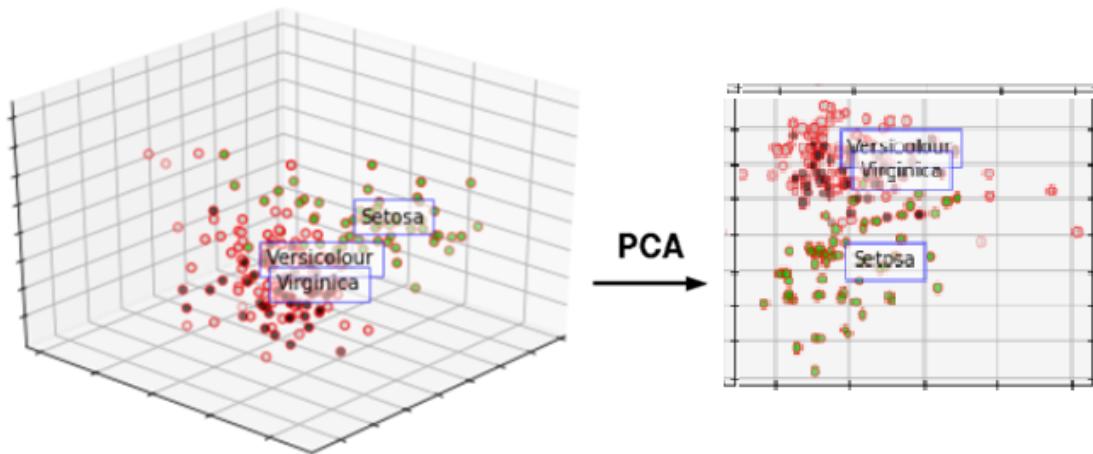


Figura 5.16: Ejemplo de la reducción dimensional por PCA de 3 características a 2 componentes principales

Hiperplano

Un hiperplano es cualquier subespacio vectorial de codimensión -1 de un espacio vectorial. De manera equivalente, un hiperplano V en un espacio vectorial W es cualquier subespacio tal que W/V es unidimensional. En términos generales, si V es un espacio vectorial n -dimensional, entonces H es un subespacio $(n-1)$ dimensional.

Un hiperplano es el núcleo de transformación lineal de cualquier mapa lineal distinto de cero desde el espacio vectorial hasta el campo subyacente.

Capítulo 6

Desarrollo experimental

Propuesta de solución

La propuesta de solución para el problema planteado consiste en una metodología compuesta por tres fases principales:

Ingeniería de Datos

En primera instancia se busca estructurar una base de datos con registros de ECG que logre unificar registros de pacientes saludables así como registros etiquetados con las patologías cardíacas de interés con el fin de lograr un Dataset enriquecido con la información y métricas necesarias para su uso en un modelo de ML. Para ello se hace uso de las bases de datos como *PTB Diagnostic ECG Database*, *MIT-BIH Atrial Fibrillation Database*, *BIDMC Congestive Heart Failure Database*, *Long Term AF Database*, *Normal Sinus Rhythm RR Interval Database*, disponibles en el sitio <https://physionet.org/> de todas ellas se tomarán a los pacientes que contengan registros de ECG y RR con una longitud mínima de 3,000 datos (una duración mínima de 5 horas 20 minutos) para garantizar que al realizar los análisis pertinentes incremente la posibilidad de analizar la evolución de las métricas propuestas en la mayor ventana de observación de tiempo posible. Se precisan en esta etapa también, las características de la base de datos MIMIC-III, la cual una vez controlados todos los procedimientos de la etapa posterior a ésta, será utilizada para la extracción de las métricas propuestas sobre una muestra de registros considerablemente mayor a la obtenida de manera sintética; de esta forma se pretende obtener de manera análoga un Dataset mucho más grande que pueda ser alimentado a un modelo supervisado de ML para clasificación binaria.

Procesamiento y análisis de señales

Una vez estructurados los datos, se procederá a leer cada registro, y realizar un pre-procesamiento a las señales que lo requieran para posteriormente obtener la señal de HRV derivada de las señales originales para su estudio. Se dividirá el análisis de las señales en sus etapas lineal y no lineal. A lo largo de ambas etapas se buscará conocer la evolución temporal producto de la aplicación de técnicas lineales y no lineales respectivamente mediante un ventaneo dinámico sobre la señal; esto permitirá derivar series de tiempo de dichas métricas sobre las cuales se podrán hacer algunas observaciones así como extraer propiedades y características tanto de la señal, así como de su distribución estadística que permitan hacer inferencias preliminares acerca de la diferenciabilidad de los estados fisiológicos que representa cada registro. Al concluir los estudios de la señal de R-R se añadirán a los Dataset creados en la fase anterior todas aquellas métricas que representan una característica notable de la señal con el afán de posibilitar su clasificación en la próxima fase.

Clasificación

Contando con estructuras de datos adecuadas para su uso, se procede en una primera instancia a resumir la información dada por cada métrica en valores puntuales interpretables por los modelos de clasificación y a realizar observaciones con respecto a la dispersión inicial de pares seleccionados de métricas derivadas de la fase anterior con el fin de realizar inferencias basadas en las propiedades intrínsecas de las mismas. Este procedimiento se realizará de manera inicial con el grupo de control, dado que es a priori conocido el hecho de que la tarea de clasificación en el grupo de control contra un grupo de registros patológicos es relativamente sencilla y no representará un reto para el modelo de aprendizaje propuesto. Se realiza en segunda instancia el procedimiento mencionado ahora con el fin de explorar y evaluar la aptitud de las métricas de las que se pueda notar algún comportamiento que suponga una posibilidad de diferenciar entre los grupos de estados patológicos.

Tomando en cuenta la información acerca de los datos que fue recabada en la etapa anterior, se procede a realizar una reducción dimensional a los datos por medio de un PCA para extraer componentes de los mismos que contengan las características más importantes de cada clase y se explora el contenido de dichos componentes. Por medio de una reestructuración sencilla se utilizan los componentes principales de cada característica para entrenar al modelo de clasificación binaria, evaluando la precisión terminal de su predicción para discriminar entre componentes con base en la información que aportan para lograr la clasificación; esto en busca de mejorar la precisión del modelo de ML.

Finalmente, se realiza un estudio del rendimiento del clasificador tomando su evaluación como antecedente para discutir el reconocimiento de patrones efectuado en el trabajo y todas las consideraciones necesarias, bajo la definición conceptual, para la identificación de la existencia de precursores o biomarcadores dinámicos que describan un posible mecanismo de cambio de estado fisiológico saludable a uno patológico.

Ingeniería de Datos

DUMMY

Dado que el número de registros contenidos en la base de datos MIMIC-III, así como el acceso controlado a la misma representó un reto en diversos aspectos en etapas tempranas del desarrollo de este proyecto. La concatenación de una base de datos de prueba se hizo necesaria para realizar análisis iniciales que de otra forma hechos sobre MIMIC-III incrementarían el tiempo de ejecución de los algoritmos ya que de requerir ajustes sería necesario esperar a que el algoritmo recorra toda la base de datos cada vez que se ejecute ó bien, representaría una pérdida de tiempo al fallar en algún registro que se encuentre pasada la mitad de la base de datos.

Es así que para probar el funcionamiento de los diversos análisis y facilitar la creación de funciones que realicen posteriormente toda clase de trabajo sobre MIMIC-III se hace uso de un **Dataframe Unificador de Métricas, Marcadores Y registros**, identificado para mayor comodidad y facilidad de lectura con el acrónimo **DUMMY**.

Bases de datos consideradas en DUMMY

Para estructurar el Dataset de estudio se recopilaron registros de las bases de datos que se detallarán a continuación de las cuales se extrajeron los casos saludables ó de control y aquellos casos marcados con las patologías de interés exceptuando registros de pacientes etiquetados con IAM debido a que los casos recopilados de las distintas bases de datos fueron excluidos al presentar una duración máxima de 2 minutos.

La siguiente lista muestra una descripción detallada de las características de las bases de datos utilizadas para la formación de DUMMY. Cada registro será guardado dentro de la instancia dedicada en DUMMY una vez preprocesada la señal y convertida en una señal de RR. Una muestra ilustrativa de los registros sin alteraciones previas puede ser encontrada en el Apéndice 10:

- **MIT-BIH Atrial Fibrillation DB:** Base de datos de 25 registros largos de sujetos de prueba humanos con fibrilación atrial. Con un registro promedio de 10 horas de duración con dos canales de registro.
 - Frecuencia de muestreo: 250Hz
 - Resolución: 12 bits
- **BIDMC Congestive Heart Failure DB:** Base de datos con 15 sujetos de prueba humanos con caso severo de insuficiencia cardíaca congestiva, los registros presentan una duración promedio de 10 horas, con un registro de 2 canales.
 - Frecuencia de muestreo: 250Hz
 - Resolución: 12 bits
- **MIT-BIH Arrhythmia DB:** Contiene 47 registros de media hora de registros ambulatorios con dos canales de ECG, obtenidos de sujetos de prueba estudiados por el Laboratorio de Arritmia del Boston Israel Hospital entre 1975 y 1979.
 - Frecuencia de muestreo: 360Hz
 - Resolución: 11 bits
- **Long Term AF DB:** En esta base de datos se presentan 84 registros a largo plazo de ECG, de sujetos de prueba con fibrilación atrial sostenida o paroxística. La duración del registro de los casos varía pero se encuentran con un promedio de 24-25 horas.
 - Frecuencia de muestreo: 128Hz
 - Resolución: 12 bits
- **PTB Diagnostic ECG DB:** 549 Registros obtenidos de un equipo prototípico con 16 canales de entrada. La base de datos abarca 290 sujetos de prueba humanos (de 17 a 97 años, 209 hombres, 81 mujeres) con diversas condiciones, de las cuales se enfocará en los 148 casos de Infarto al Miocardio, y los 52 sujetos de control.
 - Frecuencia de muestreo: 1000Hz
 - Resolución: 16 bits
 - Rango: + – 16,384mV

MIMIC-III

La Medical Information Mart for Intensive Care - III (MIMIC-III) es una base de datos que incluye información relacionada con pacientes ingresados a una unidad de cuidado intensivo. Los datos incluyen señales vitales, medicamentos, pruebas de laboratorio, observaciones y notas realizadas por personal médico del hospital, códigos de procedimientos, reportes de imágenes y duración de hospitalización. Esta base soporta aplicaciones para investigaciones industriales y académicas, mejoras de calidad y educación.

Comprende 62532 registros de pacientes internados, es completamente gratuita y su acceso se encuentra sujeto a la acreditación del curso de Investigación única de datos o especímenes elaborado por el Massachusetts Institute of Technology.

Lista de Tablas

La base de datos contiene 40 tablas en las cuales cada columna desglosa la información del paciente, estas se conectan por medio de un identificador ID único para cada sujeto.

Las tablas que son usadas para definir y rastrear la estancia de los pacientes son las siguientes:

- ADMISSIONS. Cada hospitalización para cada paciente en la base de datos
- CALLOUT. Información correspondiente a la alta de cuidados intensivos
- ICUSTAYS. Cada estadía en la UCI
- PATIENTS. Cada paciente único en la base de datos
- SERVICES. El servicio hospitalario mediante el cual queda registrado el paciente
- TRANSFER. Movimientos de cama a cama del paciente dentro del hospital

Las siguientes tablas contienen datos registrados en la estancia del paciente en la UCI:

- CAREGIVERS. Cada cuidador que registro información a la base
- CHARTEVENTS. Cada observación para cada paciente
- DATETIMEEVENTS. Todos los registros de observación en fechas
- INPUTEVENTSCSV. Registros de pacientes monitoreados usando el Philips CareVue system dentro de la UCI
- INPUTEVENTSMV. Registros de pacientes monitoreados usando el iMDSoft Metavision dentro de la UCI
- NOTEVENTS. Notas del paciente por personal de enfermeria asi como reportes de ECG,imagenes,etc.
- OUTPUTEVENTS. Información de salida de los pacientes mientras se encuentran en la UCI
- PROCEDUREEVENTSMV. Procedimientos para pacientes monitoreados en la UCI utilizando el iMDSoft MetaVision system.

Las siguientes tablas contienen datos registrados en el archivo del hospital:

- CPTEVENTS. Procedimientos registrados como Terminología procesal actual
- DIAGNOSESICD. Diagnósticos hospitalarios utilizando el sistema ICD
- DRGCODES. Grupos de relativos de diagnóstico
- LABEVENTS. Pruebas de laboratorio para pacientes con estancia en UCI y de consulta externa.
- MICROBIOLOGYEVENTS. Pruebas de microbiología
- PRESCRIPTIONS. Medicamentos preescritos pero no necesariamente administrados
- PROCEDURESICD. Procedimientos efectuados en pacientes utilizando el sistema ICD.

Por último las siguientes tablas son diccionarios

- DCPT. Diccionario de Alto nivel de terminología procesal actual
- DICDDIAGNOSES. Diccionario de la clasificación estadística internacional de enfermedades y problemas de salud relacionados
- DICDPROCEDURES. Diccionario de códigos de la Clasificación Estadística Internacional de Enfermedades y Problemas Relacionados con la Salud (CIE) relacionados con los procedimientos
- DITEMS. Diccionario de ITEMIDs que aparecen en la base de datos MIMIC, excepto aquellos que se relacionen con pruebas de laboratorio
- DLABITEMS. Diccionario de ITEMID en la base de datos del laboratorio relacionados con las pruebas de laboratorio

Una representación esquemática de la base de datos MIMIC-III, todos sus componentes y la relación que existe entre ellos puede verse en la figura 6.1.

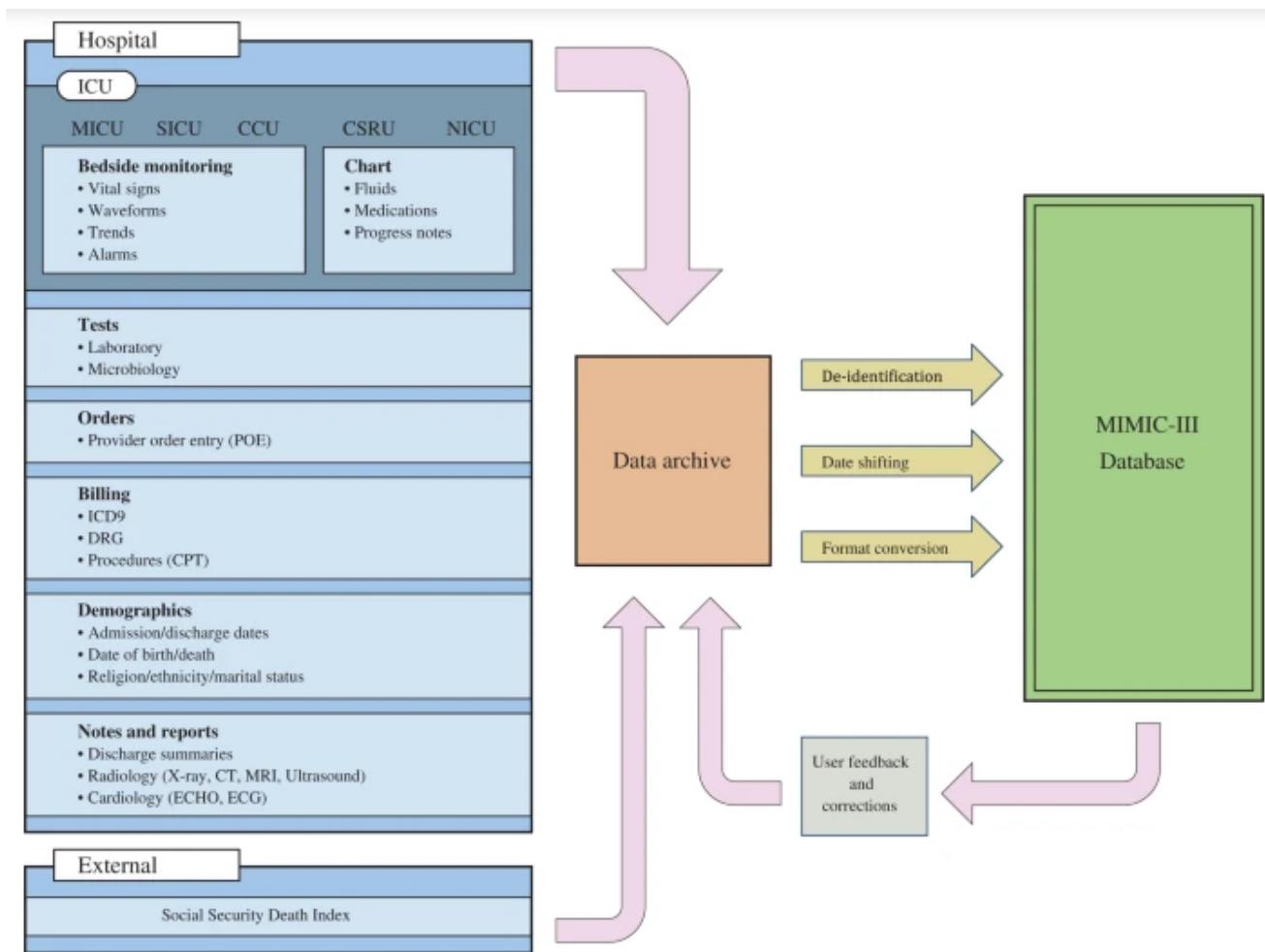


Figura 6.1: Descripción general de la base de datos de cuidados intensivos MIMIC-III.

MIMIC-III Waveform Database Matched Subset

MIMIC-III (Medical Information Mart for Intensive Care III) tiene un compendio de archivos de señales elecardiográficas y otros signos vitales en *"The MIMIC-III Waveform Database"*, sin embargo pesa 6.7TB en su totalidad, y dichas señales no tienen diagnóstico relacionado directamente. Para solventar esta situación existe la base de datos *"MIMIC-III Waveform Database Matched Subset"*, 2.4TB de señales emparejadas con su diagnóstico correspondiente. De dicha base emparejada se realizó la búsqueda de las patologías correspondientes a éste estudio y se encontraron el siguiente número de registros de ECG para pacientes ingresados en la Unidad de Cuidado Intensivo:

- Infarto agudo de miocardio - 6109 Registros
- Fibrilación Atrial - 1052 Registros
- Insuficiencia Cardíaca Congestiva - 9591 Registros

Contando en su totalidad con 14865 señales de ECG las cuales tienen las siguientes características demográficas y las cuales se ilustran en las Figuras 6.2 y 6.3:

Por sexo:

- Sexo Masculino - 1842 Pacientes
- Sexo Femenino - 1446 Pacientes

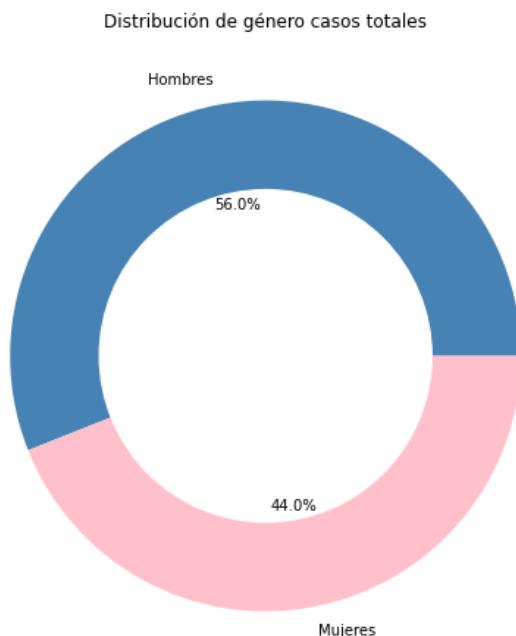


Figura 6.2: Distribución de género de los registros extraídos de MIMIC-III

Por grupo étnico:

- Blancos - 2231 Pacientes
- Desconocido - 453 Pacientes
- Afroamericano - 312 Pacientes
- Latinos - 61 Pacientes
- Asiáticos - 21 Pacientes

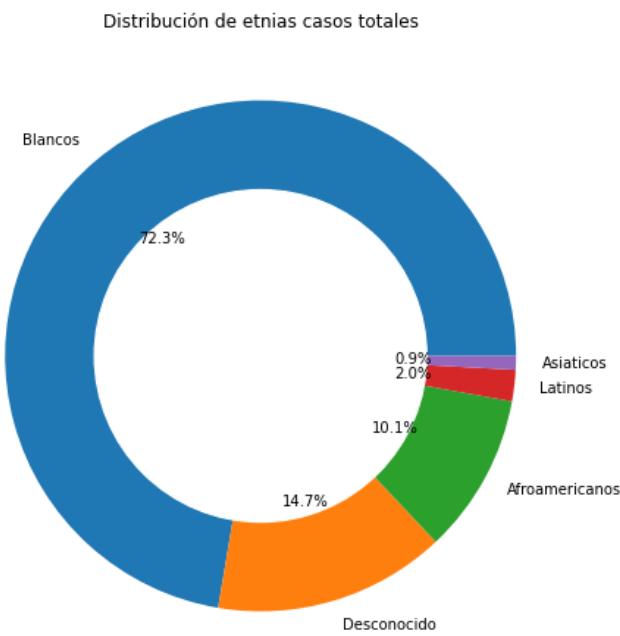


Figura 6.3: Distribución de grupos étnicos de los registros extraídos de MIMIC-III

Los registros de ECG se encuentran de manera específica en MIMIC-III WaveForm Database Matched Subset, por lo que fue necesario primero identificar el diagnóstico en la base de datos principal, extrayendo así el id de paciente para encontrar posteriormente los registros pertenecientes a ese identificador. Al ser una base de datos de pacientes en UCIs, esta no cuenta con registros de pacientes sanos que formen el cohorte para control. Por esta razón, se incluirá dicho grupo de DUMMY para completar así la estructura de Dataset a clasificar.

Descarga de bases de datos

Las bases de datos de prueba no presentaban un tamaño obstrusivo ni dificultades de descarga, la plataforma physionet.org propicia los métodos para el acceso total a los datos, sin embargo para descargar la sección con datos de interés de MIMIC-III Matched Waveform Data Base (Base de datos de señales emparejadas), es decir el conjunto de registros cuyos pacientes diagnosticados tuvieran FA, ICC e IAM, se tomó en cuenta que requeriría un tiempo extenso para descargar incluso una sección menor de los 6.7 TB de datos totales, con ésto se elaboró un código ejecutable en BASH que pudo ser cargado a un servidor remoto para obtener todos los casos y sus respectivos registros. Inicialmente, con el acceso oficial a MIMIC III se obtuvo el listado de pacientes con sus respectivos diagnósticos, lo que permitió generar un filtrado y solo identificar las claves de identidad de casos relevantes, a continuación una representación del flujo de trabajo del código desarrollado para la obtención oficial de datos en la Figura [6.4].

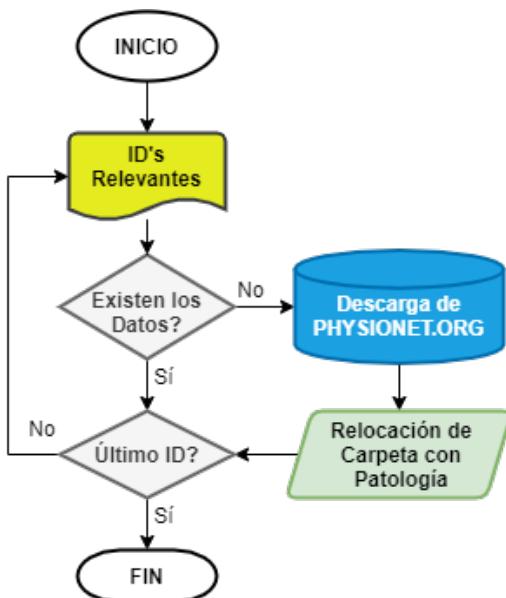


Figura 6.4: Diagrama de Flujo para descarga de MIMIC III

Lectura de Datos y estructuración de Datasets

Los datos descargados tanto de las bases de prueba así como de MIMIC-III continúan haciendo uso de el mismo tipo de archivos comprimidos para registrar las señales ECG, la diferencia principal radica en que debido a la extensión de MIMIC-III dichos archivos comprimidos fueron agrupados de manera diferente, por lo que se prosiguió a desarrollar dos metodologías de lectura para cada una.

Reducción de las bases

En el caso de DUMMY, donde se dispone de los archivos comprimidos en una misma ubicación, la lectura fue más directa, sin embargo en el caso de MIMIC-III los registros a su vez se encuentran agrupados en carpetas que corresponden a un mismo sujeto, es decir que cada sujeto de ésta base de datos cuenta con múltiples registros de ECG en lugar de sólo uno como en la base anterior, haciendo uso de los datos extras provistos para indexar los registros internos en cada carpeta se desarrolló un código con el paradigma de Programación Orientada a Objetos para optimizar la forma en que cada registro fuera revisado, anexado y compactado mientras no se esté procesando, lo que se vio reflejado en tiempos de procesamiento menores y una mayor facilidad de manipulación de todos los casos.

Preparación del Dataset para el pre-procesamiento

Como preámbulo a la fase siguiente es necesario preparar la base de datos para el pre-procesamiento de la señal de ECG. Para lograr esto se aplicaron también dos metodologías aunque se emplearon las mismas herramientas de detección. Para la base DUMMY (Figura 6.5) se realizó en varias etapas; inicialmente, durante la lectura, debido a que contienen diversas cantidades de señales registradas. Bajo una inspección visual de cada registro se logró identificar aquellos donde las pruebas de detección de picos tienen un mejor funcionamiento. Una vez seleccionada la señal principal a emplear ésta puede ahora ser introducida en la función de detector de picos que se utilizará en el pre-procesamiento de la fase posterior. Con relación a las instancias de los Datasets iniciales se almacenaron los datos, características, diagnóstico y comentarios originales a cada una de las bases de datos seleccionadas y fueron almacenados en una serie de archivos tipo “json”, esto debido a que este formato hace más eficiente la manipulación y ahorra el proceso de detección de R-R en cada ocasión que fuera este requerido.

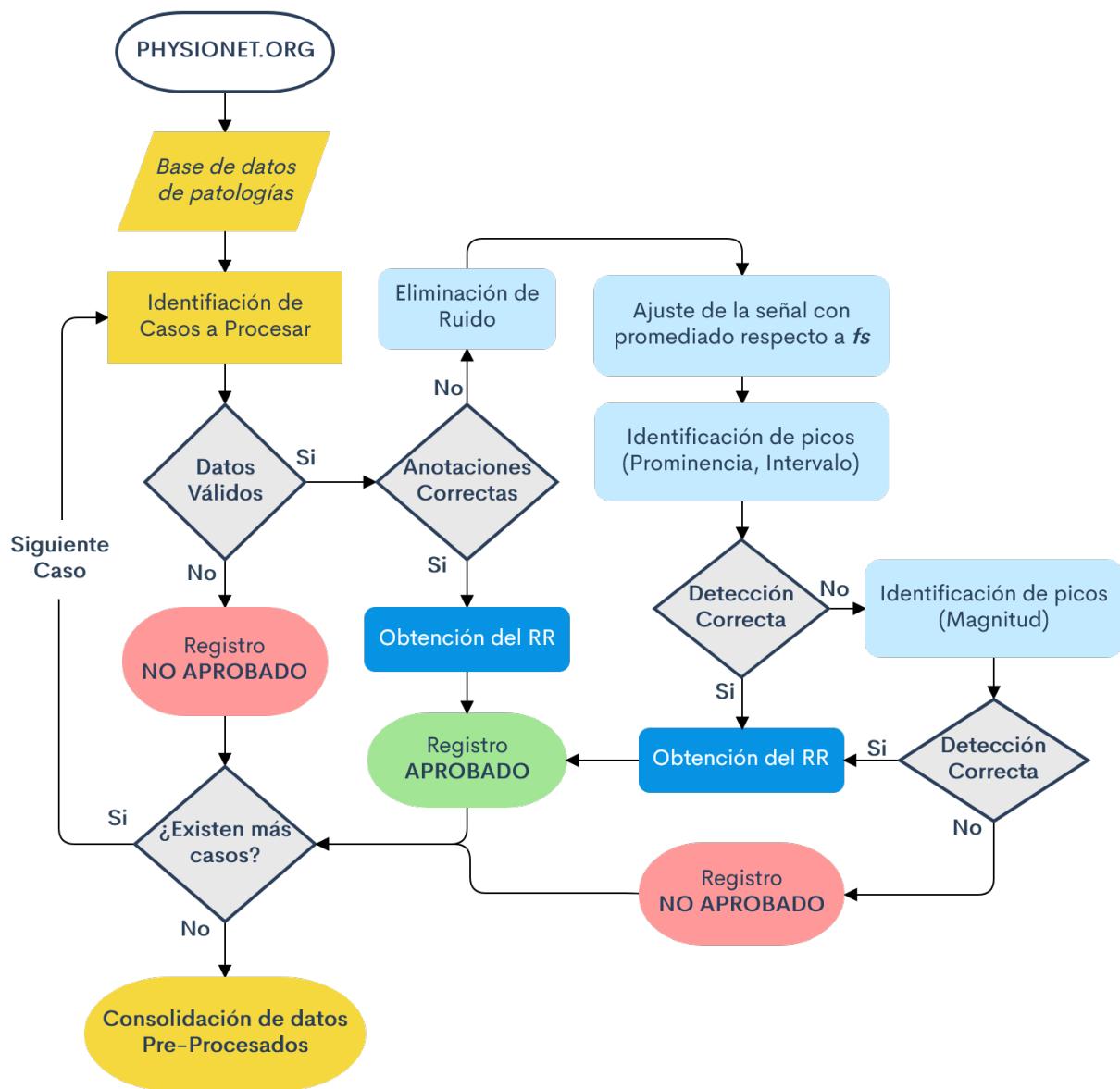


Figura 6.5: Diagrama de flujo del pre-procesamiento para bases de prueba

La manipulación de los datos de MIMIC-III requirió de la generación de clases para Casos y Registros. Esto para permitir identificar puntualmente a cada uno de los registros patológicos de cada sujeto sin separarlos realmente entre sí y mezclarlos con otros. El procedimiento, ejemplificado por el diagrama de la figura 6.6, fue más extenso debido a estos ajustes; sin embargo permitió integrar directamente las diferentes etapas del proceso. Para la identificación de la carpeta contenedora de los registros (identificada por el id del sujeto) se realizó la lectura de todos los documentos internos listados dejando un registro en el código con la información más importante en el procedimiento de carga para futuras referencias. Para mantener la consistencia de los datos se encontró la forma de señal con mayor presencia en cada caso y se extrajeron de todos ellos, los registros que no la tuviesen, o bien los que contuvieran series de tiempo de longitud inconsistente ó datos médicos puntuales no relacionadas directamente con la dinámica cardiaca (p.ej. Respiración, Temperatura, etc...) fueron excluidas desde esta etapa de filtrado final del Dataset. Como paso final se realizó la prueba para la generación de la señal R-R a partir del ECG original de cada registro, permitiendo la obtención del vector R-R sobre el cual se basa el análisis en este trabajo.

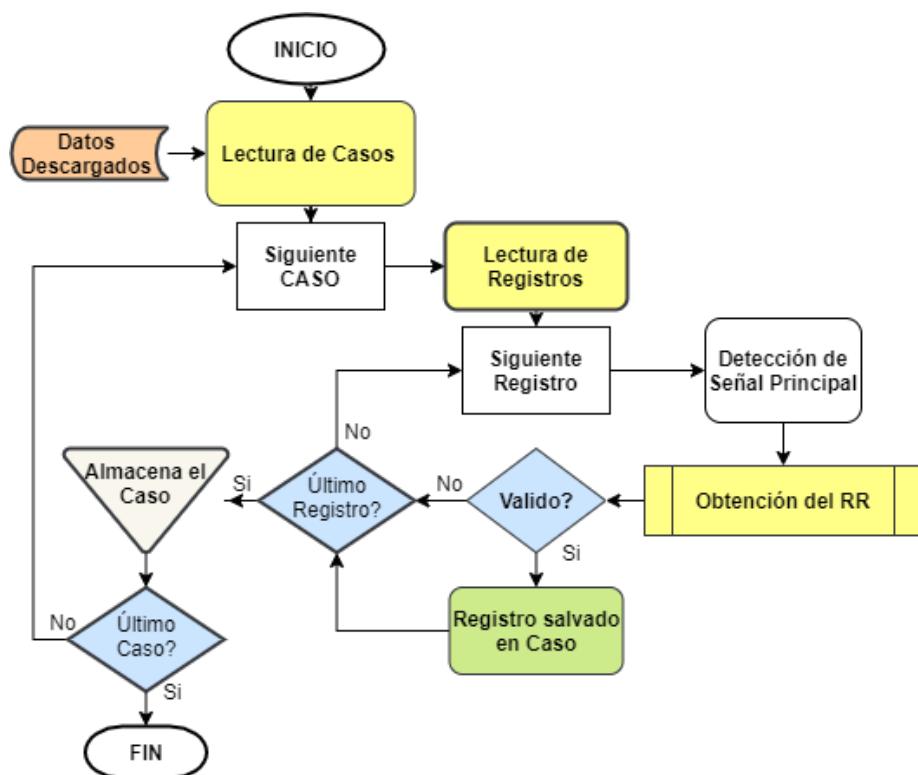


Figura 6.6: Diagrama de flujo del Pre-procesamiento para MIMIC

Procesamiento y análisis de señales

Pre-procesamiento de la señal ECG

Lectura de señales ECG

Una vez descargadas las bases de datos y preparada para la recepción de la señal se obtuvo el acceso a los archivos directos de cada registro, para poder hacer uso de éstos se empleó el módulo de python “wfdb” [49], el cual permite un acceso rápido a las notas y detalles de las señales. Siguiendo la estructura del diagrama de flujo, se identifican los casos a procesar condicionando al proceso a que en caso de que no existan datos de señales de ECG, se registrara en la bitácora del proceso un sujeto no aprobado y pasará al siguiente caso.

Si existen señales válidas en el caso, es posible que el registro cuente con las anotaciones pertinentes en los picos de la señal 6.7, corroborados por el personal médico que consolidó la base de datos, en caso de ser así se extraen esas coordenadas y se genera la señal de variabilidad cardíaca, registrando el caso con sus características y marcándolo como aprobado. Sin embargo también existen bases de datos que no cuentan con archivo de anotaciones, o bien dichas anotaciones se han registrado de forma incorrecta, esto está indicado en las bases de datos. Para poder hacer uso de éstas señales, en la fase de revisión de anotaciones (Figura 6.6) cuando es considerado un mal registro, se procede a realizar un procesamiento alternativo.

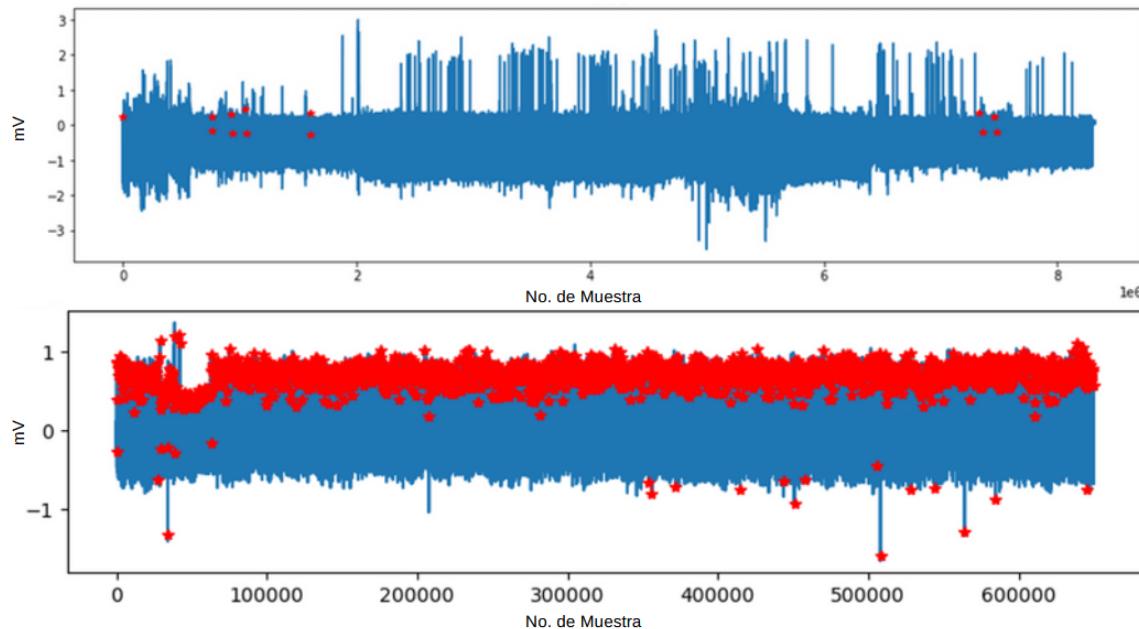


Figura 6.7: Señal con anotaciones correctas(abajo) e incorrectas(arriba)

Reducción de ruido en la señal

El primer paso encaminado a la obtención de la señal de HRV es remover en la mayor medida posible ruidos y componentes cíclicas en la señal. La forma de realizar el primer tratamiento es por medio de un filtro Notch de Respuesta Infinita al Impulso de segundo orden (“Second order IIR Notch filter” en inglés)[50] a través de la librería “scipy”. El filtro aplicado tiene una frecuencia de corte de 60Hz con un factor de calidad Q de 1. En la figura 6.8 encontramos la misma señal que anteriormente carecía de anotaciones en la figura 6.7.

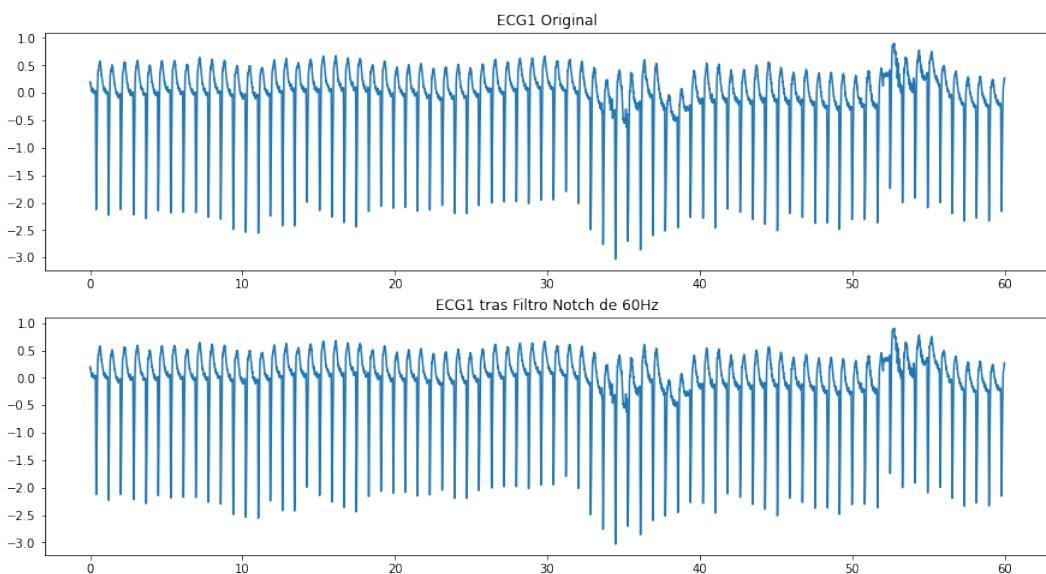


Figura 6.8: Señal antes(Arriba) y después (abajo) del filtro.

Corrección de Línea base errante

Debido a la metodología por la cual se registran los ECG, es imposible detectar exclusivamente los cambios de potencial ocurriendo en el corazón si los electrodos que se encargan de ello se encuentran en la superficie de la dermis. Es por ello que es posible percibir movimientos en la señal que podrían corresponder al movimiento del sujeto, a su respiración, etc. Esto dificulta en cierta medida la detección de los pulsos a lo largo de la señal, por lo que llega a ser de especial importancia llegar a cancelar en la medida de lo posible estas variaciones.

Para poder lograr esto se aplica un filtrado de media móvil en dos fases; para determinar cuantos datos deberá abarcar en cada caso para registrar su promedio se tomó en cuenta la frecuencia de muestreo fs y un valor pulsos por minuto como límite superior de 230 BPM, éste valor fue propuesto tomando en cuenta la estimación más alta de un individuo realizando un esfuerzo máximo a sus 20 años, de 200 BPM, y agregando 30 más para considerar un límite insuperable. Así pues la ventana del filtro fue determinada de la siguiente manera:

$$Ventana = \frac{60\text{seg} \cdot fs}{\text{Max BPM}}$$

Tomando este tamaño de ventana (W) con una señal S de n datos se aplica un promedio(\bar{a}) de cada ventana de tal forma que:

$$\bar{a} = \frac{1}{W} \sum_{i=0}^{n-1} S_{W-i}$$

Obteniendo la señal de media móvil directa se encuentran movimientos respecto a pequeños y grandes cambios, pero para registrar exclusivamente los cambios mayores, se realiza nuevamente el filtro de media móvil respecto al filtro anterior con el doble de ventana por 5 iteraciones, el resultado se puede observar en la parte superior de la figura 6.9.

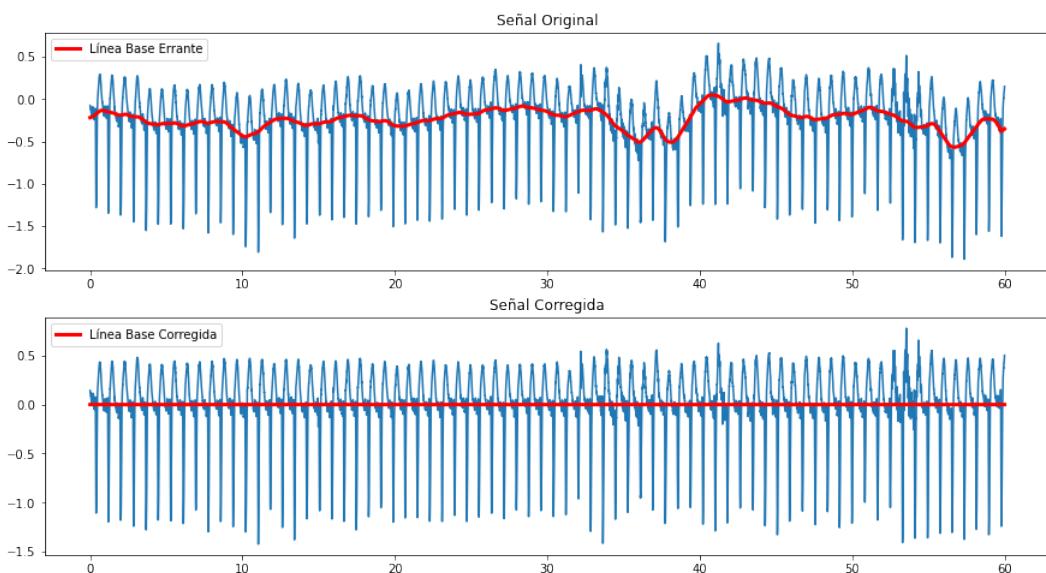


Figura 6.9: Media móvil generada(superior) y Señal con el ajuste de línea(inferior)

Detección de picos y obtención de señal R-R

Aplicando la función de detección de picos del módulo “scipy”, se establecen parámetros para determinar los picos que corresponden a la onda R; el primero de ellos es que sean datos que presenten una prominencia con respecto a puntos en su alrededor a partir de un 80 % y que exista un distanciamiento entre dichos puntos de por lo menos el tamaño de ventana w definido anteriormente con un BPM de 230, como se ve en la muestra superior de la figura 6.10. De esta manera al conocer las coordenadas de cada punto detectado, es posible calcular la separación entre ondas R y cuantificar dicha distancia del interlatido cardiaco para generar un vector de variabilidad cardíaca (HRV o R-R) tal y como se demuestra en el cuadro inferior de la figura 6.10.

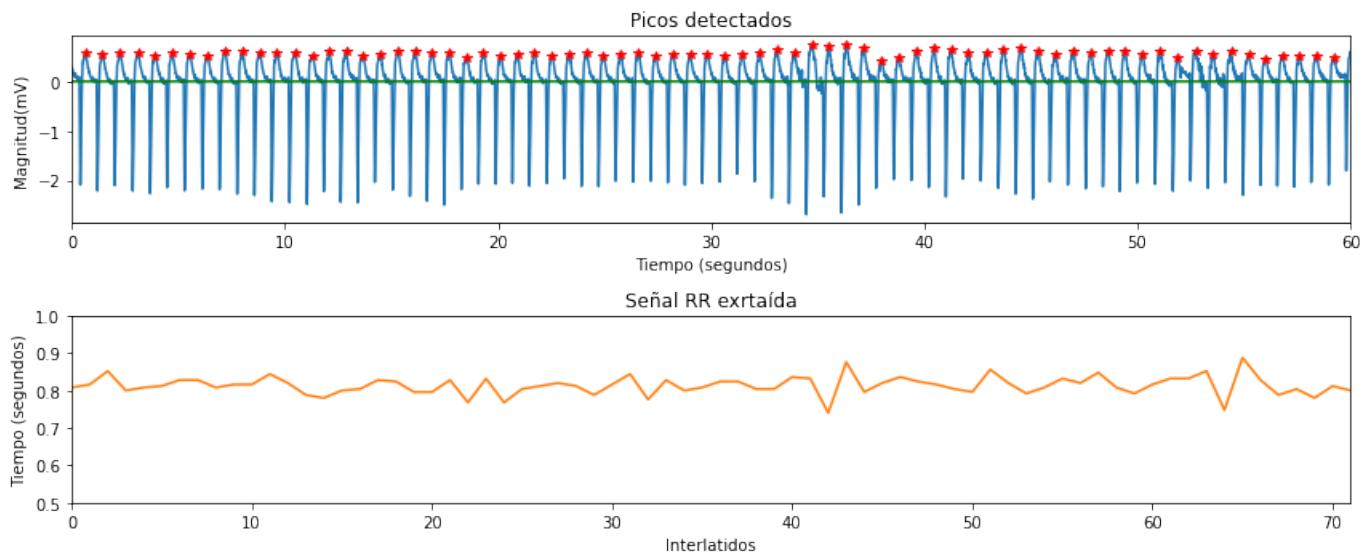


Figura 6.10: Detección de picos y señal de variabilidad cardíaca obtenida.

Un análisis de HRV de 24 horas nominales es considerado una observación de mediano plazo, mientras un que corto plazo se sitúa en una ventana de 5 minutos y cualquier observación por debajo de ese margen será considerada de ultra-corto plazo. Esta precisión es en muchos casos determinante para obtener características de la señal de HRV, ya sea dentro del dominio del tiempo, de la frecuencia ó características propias de un análisis no lineal, debido a sus complejidad.

Para notar si la evolución temporal de una característica del R-R se encuentra dentro de la normalidad ó comienza a desviarse a valores anómalos, es necesario realizar la observación dentro de una ventana de tiempo suficientemente grande como para descartar que una fluctuación en el valor de dicha característica sea un evento ectópico y presente un patrón recurrente y reconocible con tendencia a reproducirse a largo plazo.

Es por ello que se busca conocer este intervalo donde los valores de las características se consideran normales; de esta manera, pueden ser tomados como una referencia para conocer el comportamiento esperado de la señal RR y lograr identificar si dicha señal posee características que van paulatinamente saliendo del estado saludable.

Metodología para el procesamiento de la señal R-R

Con todos los vectores R-R obtenidos y antes de ser procesados por los métodos lineales y no-lineales se aplicó un último filtro. Realizando varias pruebas de eficiencia se optó por emplear una ventana de 500 datos (es decir, 500 interlatidos) para aplicarle funciones matemáticas, dicho esto, todos los registros con una longitud menor a seis veces el tamaño de ventana fueron omitidos también.

Técnica de ventaneo dinámico

Una vez que se cuenta con las señales R-R derivadas del procedimiento anterior, se procede a la metodología exploratoria de las señales mediante la aplicación de una ventana deslizante a cada señal de interlatidos.

En este sentido, el tamaño de ventana juega un papel fundamental en el cálculo de cualquier característica de la señal. Abrir una ventana de observación de 24 horas en una señal de HRV mostraría un espacio en el que es más probable que ocurran eventos anómalos en comparación con ventanas más cortas, sin embargo el cálculo de métricas como son los momentos estadísticos de esta ventana pueden generalizar el comportamiento de la señal e ignorar fenómenos latentes o localidades específicas de la señal. En el caso contrario se pensaría entonces que al tomar una observación de ultra-corto plazo y deslizar a través del tiempo capturaría por completo todos esos eventos locales sin ignorar información, sin embargo debido a la alta dimensionalidad (comportamiento fractal) del HRV esto podría causar problemas al tratarse de un evento autosimilar libre de escala y terminar por saturar la estimación de las métricas dejando una medición francamente inservible.

Tomando en consideración este hecho, el análisis a realizar sobre la señal contemplará una ventana de observación de corto plazo (Aproximadamente 8 minutos) sobre la variabilidad cardiaca, que al deslizar a lo largo de una señal prolongada amortiguaría el efecto anteriormente descrito para darle mayor significado a la evolución en el tiempo de las métricas calculadas. Con estas consideraciones, se diseña una ventana con las siguientes características propuestas:

- *Tamaño de Ventana:* 500 datos
- *Superposición:* 0,5

De manera que, una vez obtenida la ventana que contendrá una fracción de la señal original, se realizarán los cálculos de las métricas deseadas los cuales capturan en un vector independiente, inicialmente vacío, el valor puntual correspondiente a la observación de una técnica. Al finalizar la captura de dicho valor, la ventana de observación es deslizada en sentido del crecimiento del eje temporal hasta el índice determinado por el valor de superposición propuesto y se repetirá el procedimiento anterior hasta alcanzar el final de la longitud de la señal.

La utilidad de esta técnica se encuentra en que al deslizar esta observación a lo largo del eje temporal de la señal original los valores rescatados son equivalentes a conocer, bajo ciertas condiciones definidas por el tamaño de la muestra tomada, la evolución temporal de las métricas deseadas.

Análisis Lineal de la señal R-R

La primer etapa del trabajo consiste en la exploración de la efectividad de los momentos estadísticos de la señal y su respectivo coeficiente de variación al comparar un grupo de control de pacientes saludables con una muestra de registros tanto de la base DUMMY para su análisis y discusión y posteriormente de MIMIC-III para un aumento significativo en el tamaño de muestra. Esto con el fin de mostrar si estas características de la señal son capaces de diferenciar entre estados saludables y patológicos.

Para este estudio se tomaron los siguientes registros de las bases de datos:

- *Pacientes sanos:* 70 registros
- *Pacientes con patología cardíaca:* 195 registros

El análisis aplicado a la señal en cada registro es procesado mediante la técnica de ventaneo dinámico y genera respectivamente vectores de los momentos estadísticos (media, varianza, sesgo y curtosis) así como el coeficiente de variación, éstos vectores son los que se usarán para identificar el comportamiento de la señal y sus tendencias.

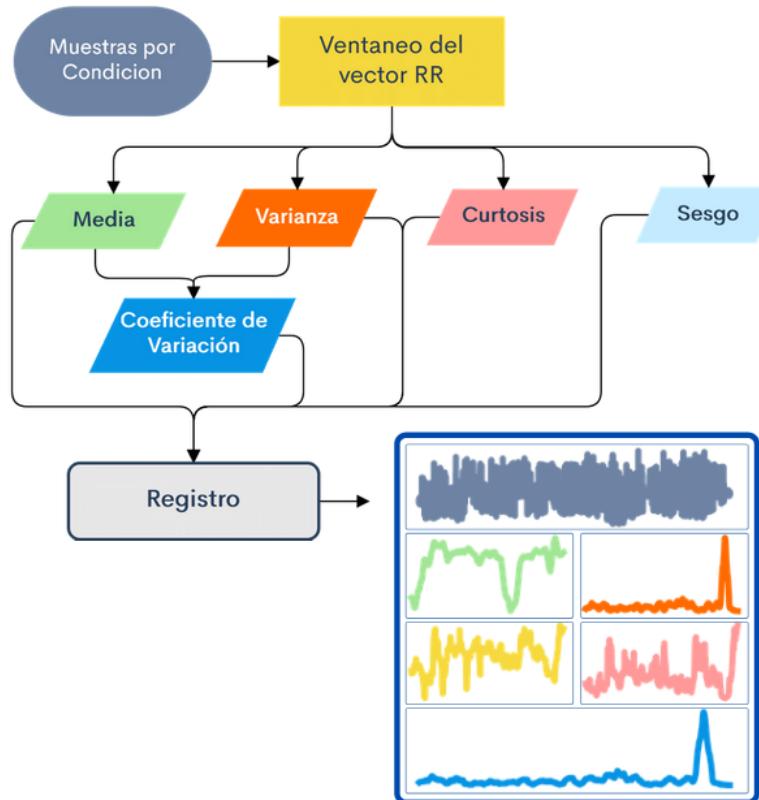


Figura 6.11: Diagrama de flujo del análisis lineal

En la figura 6.12 podemos visualizar el comportamiento temporal de las métricas lineales que comprenden a un mismo registro, en la sección superior se observa la señal rr en su totalidad para un registro saludable de la cual el ventaneo dinámico produce los vectores observados en las retículas inferiores:

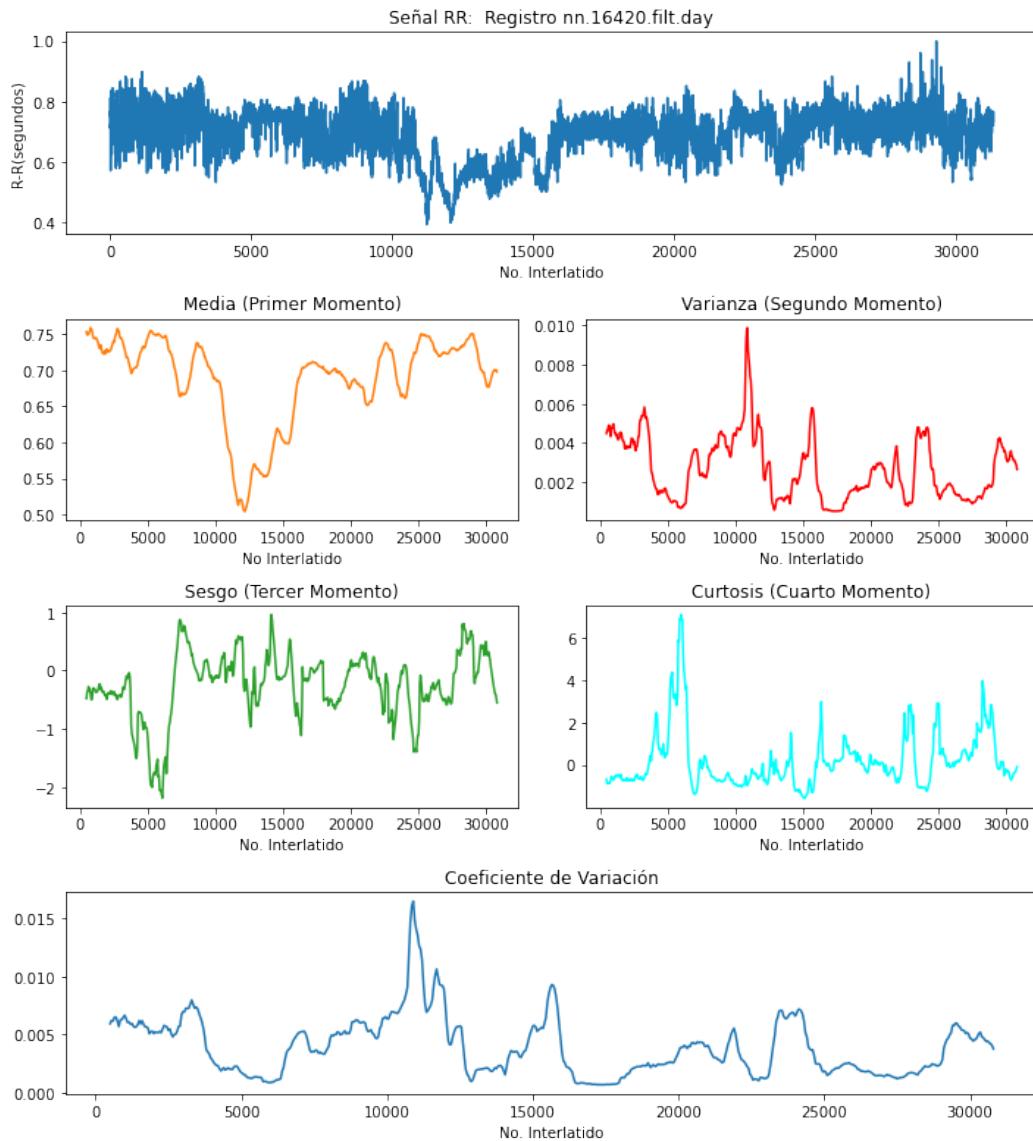


Figura 6.12: Gráfica de momentos Sano

Todas las señales RR pasaron por un proceso similar para generar sus respectivos vectores de análisis.

Visualización de momentos estadísticos

La visualización del grupo de sujetos sanos de la figura 6.13, se inspecciona para inferir si existe un patrón en su comportamiento que pueda considerarse como un punto de comparación claro para las señales patológicas.

En primer lugar se puede observar que el comportamiento del primer momento no presenta alguna característica diferenciable a lo largo de toda su evolución, no existe alguna zona en el intervalo del valor en la cual se contenga el comportamiento. La gráfica del segundo momento indica un ascenso que rebasa en algunos casos los valores de 0.02 que la mayoría de los casos parece no sobrepasar, fuera de estas zonas de crecimiento no parece tener algún patrón inmediato en su comportamiento temporal debido a que se presentan de manera inconsistente. En cuanto al valor del sesgo, este oscila cerca de 1 y experimenta un ascenso en un caso hasta rebasar un valor de 2, donde puede considerarse un sesgo negativo elevado. De manera similar, la curtosis llega a un comportamiento leptocúrtico rebasando el valor de 60 llegando a tomar valores de hasta 90 en un único caso, sin embargo la gran mayoría de los casos está contenida en un valor de 10. Finalmente se observa que el coeficiente de variación, normalmente contenido en un rango de 0.000 y 0.050 aproximadamente, también asciende a un valor máximo cerca del mismo punto en que los momentos 2, 3 y 4 sufren una variación.

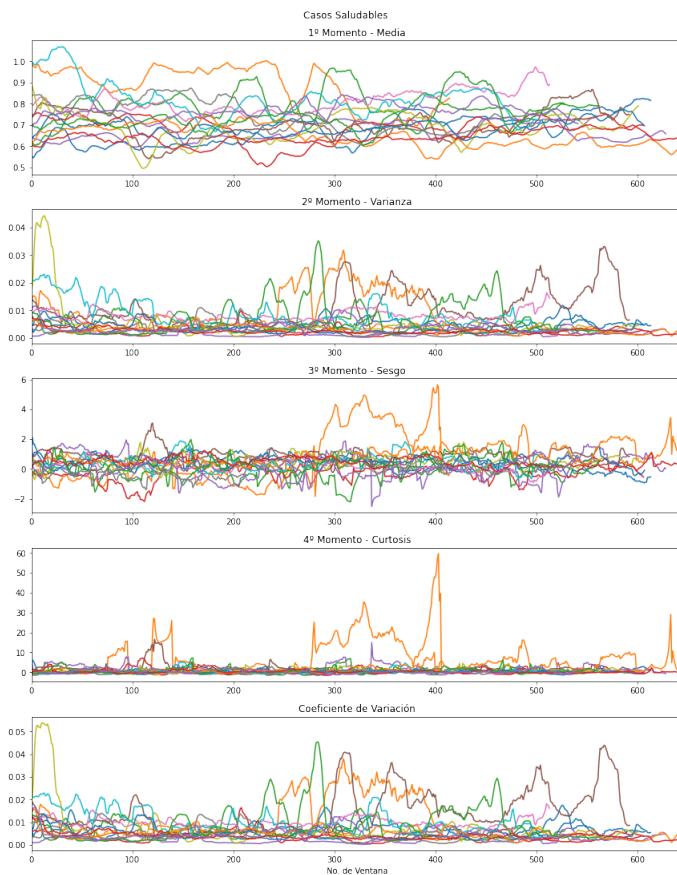


Figura 6.13: Gráfica de momentos Sano

Para las señales correspondientes a los casos patológicos, tanto para Fibrilación Atrial, como para Insuficiencia Cardiaca Congestiva se observa un rango mayor ocupado en la magnitud de la señal del interlatido cardiaco con respecto a la señal sana. Observando la gráfica correspondiente a la media de ambos casos patológicos puede notarse que oscilan alrededor de valores ligeramente distintos a la media del grupo sano; en un rango ligeramente mayor para el caso de FA con valores máximos de 0.8 y pronunciado para el caso de ICC con valores hasta 0.9.

Para ambos casos patológicos se hace evidente que la amplitud de la gráfica del segundo momento manifiesta valores mayores a los observables en el grupo sano; al ser este momento un parámetro que toma en cuenta a la varianza con respecto a la media de la señal, este hecho puede indicar una variabilidad superior. Para el tercer momento se observan en ambos casos valores que indican un sesgo positivo, en contraste con la señal sana.

La curtosis de la señal de FA se encuentra con valores más elevados a la sana sin superar el valor de 40, sin embargo para la señal de ICC este valor alcanza valores máximos de hasta 800 siendo valores que superan por dos ordenes de magnitud al caso sano. De manera similar a como ocurre en el segundo momento, el coeficiente de variación para ambos casos patológicos tiene valores mayores que el caso sano, siendo en el caso de ICC dónde esta diferencia es más notoria, con valores que tienen máximos que superan el valor de 0.20.

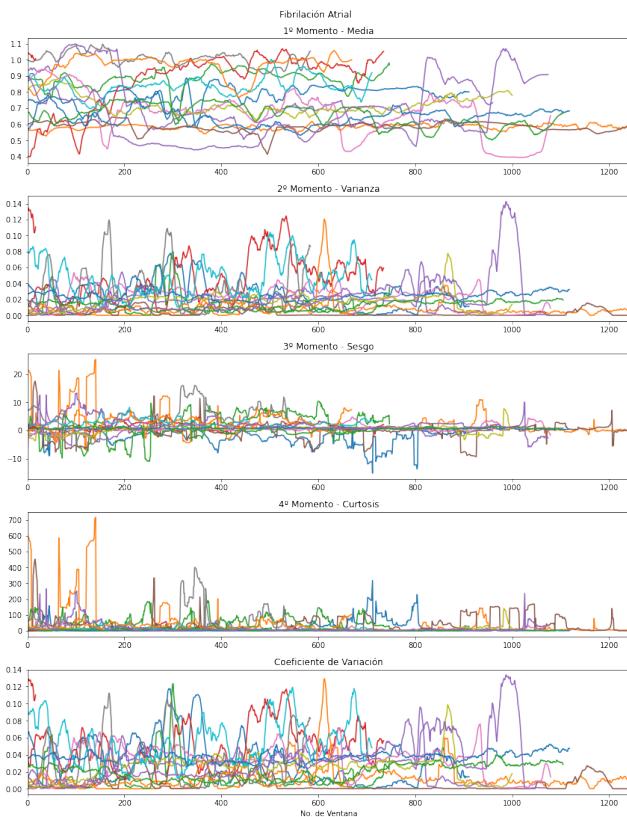


Figura 6.14: Gráfica de métricas lineales de FA



Figura 6.15: Gráfica de métricas lineales de ICC

Generación de archivos csv

Una vez procesadas las señales mediante el ventaneo dinámico, el cual genera vectores prolongados proporcionales a la longitud de la señal RR, se extraen y almacenan los valores obtenidos para media, varianza, sesgo, curtosis y coeficiente de variación en dos archivos csv; el primero con la información de las señales catalogadas como sanas y que se usarán de control, y el segundo con la información perteneciente a las señales etiquetadas como patológicas.

Lectura y limpieza de datos

Haciendo uso de la librería para python “Pandas”, estos dos archivos csv son leídos y convertidos en un objeto de tipo pandas dataframe, el cual puede ser manipulado eficientemente con el uso de funciones implementadas que facilitan consultas y operaciones. Posteriormente se implementó una etapa de limpieza de los datos para eliminar registros que tras el ventaneo y aplicación de técnicas, mostraron valores anómalos que quedaron en el registro como variables de tipo “Nan” que dificultan el análisis. Tras esta limpieza las tablas en formato dataframe obtenidas para el análisis cuentan con el siguiente formato:

Momentos estadísticos					
Registro	Media	Varianza	Sesgo	Curtosis	CV
1	m1	v1	s1	c1	cv1
2	m2	v2	s2	c2	cv2
*	m*	v*	s*	c*	cv*
N	mN	vN	sN	cN	cvN

Cuadro 6.1: Estructura de la tabla obtenida de la recopilación de las técnicas lineales

Análisis de distribución de las métricas lineales

Como primer acercamiento al análisis de la evolución temporal de cada momento, se desea conocer el panorama general de la acumulación de la muestra contenida en DUMMY mediante un histograma y su respectivo ajuste a una distribución. Estas visualizaciones ejemplifican el resultado de observar las tendencias de la distribución de frecuencia de los momentos obtenidos del intervalo R-R para cada grupo de estudio. De modo que, al obtener todas las distribuciones del grupo deseado para un momento específico se desea conocer si tienden a agruparse ó bien si existe alguna tendencia de la distribución característica que haga evidente una diferencia entre las clases.

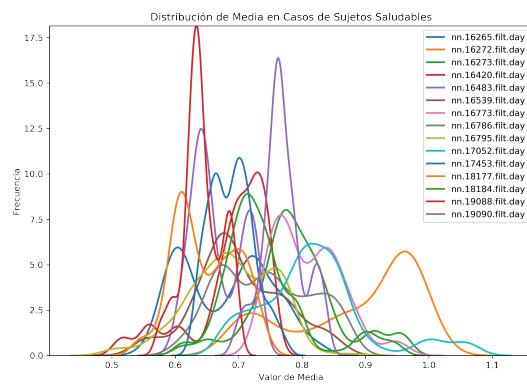


Figura 6.16: Distribución del primer comentario en casos saludables

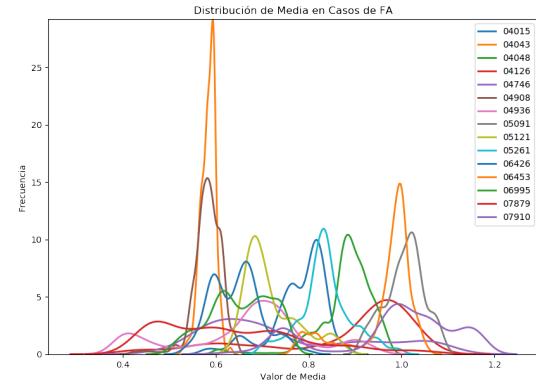


Figura 6.17: Distribución del primer comentario en casos FA

En el caso de la distribución de la media de cada grupo, cuyas características se pueden observar en el Cuadro 6.1, se encontró un rango máximo cercano entre sí para los grupos, el valor de desviación estándar fue el que presentó mayores diferencias en éste momento, sin embargo la media de los grupos permaneció en un valor cercano entre sí, ocupando el mismo espacio. Llegados a las características de la dispersión de la varianza en el Cuadro 6.2, podemos observar medias bastante similares, difiriendo sin embargo significativamente en los valores máximos y mínimos encontrados por lo que el valor de desviación estándar es casi del triple para el grupo de ECV. Para la tercera comparativa en el Cuadro 6.3, los valores máximos y mínimos de sesgo se ven superados por el grupo de enfermos, lo que afecta los valores de media y de desviación estándar significativamente. En el Cuadro 6.4 observamos la descripción estadística de la Curtosis, donde aun se presentan características en los grupos que son muy distintas entre sí, llegando el grupo de ECV a altos valores máximos de nuevo.

En las Figuras 6.18, 6.19 y 6.20 las nubes correspondientes a los grupos de estudio visualizadas en cada gráfica se presentan en las mismas áreas, lo que dificultaría su diferenciación, observando el cambio de las características representadas en los ejes es posible encontrar el comportamiento estable del grupo sano en la figura 6.21.

A partir de los resultados obtenidos en este análisis , se puede observar que partiendo de los 4 momentos estadísticos en el trabajo no puede establecerse un criterio claro que permita su diferenciabilidad. Como puede observarse en las gráficas de las Figuras 6.18, 6.19, 6.20 y 6.21, los casos saludables muestran una tendencia de media, variación y sesgo que los casos con patología cardíaca en un intervalo entre 0.6 y 1 de media, sin embargo estos valores no muestran un criterio suficiente para excluirse del otro grupo al superponerse aproximadamente con el 30 % de los casos patológicos. Con estos resultados, se observa la dificultad para la categorización de las patologías por únicamente análisis lineales que busquen discriminarse por el régimen Gaussiano, debido al limitado margen comprendido alrededor de 0.8 segundos que tienen las señales de interlatido en el espacio temporal resulta difícil encontrar una tendencia diferenciable en la presencia de patologías.

Gráficos de dispersión de los momentos estadísticos

De la misma forma se generaron gráficos de dispersión comparando en pares cada uno de los momentos utilizados en el análisis con el fin de verificar criterios de separabilidad y agrupaciones en los grupos.

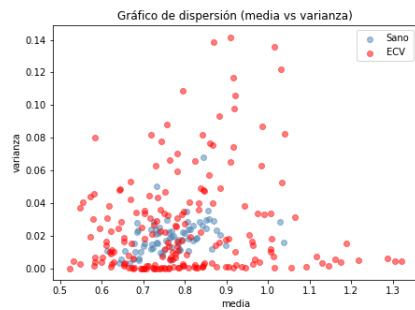


Figura 6.18: Gráfico de dispersión de la media VS varianza en los 2 grupos de estudio.

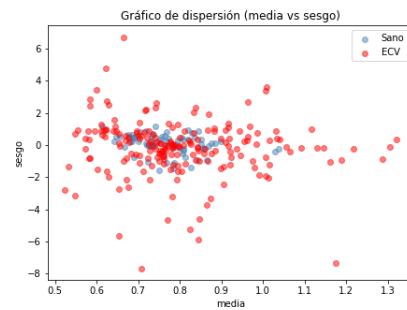


Figura 6.19: Gráfico de dispersión de la media VS sesgo en los 2 grupos de estudio.

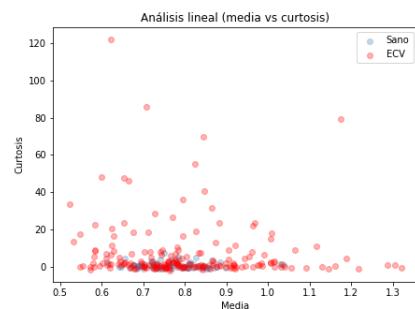


Figura 6.20: Gráfico de dispersión de la media VS curtosis en los 2 grupos de estudio.

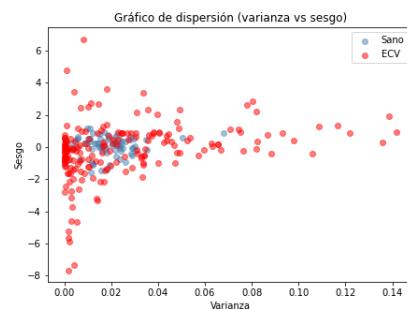


Figura 6.21: Gráfico de dispersión de la varianza VS sesgo en los 2 grupos de estudio.

En la Figura 6.18 se puede observar que los casos identificados como pacientes sanos se tienden a agrupar en un rango de media comprendido entre 0.6 y 0.9, mientras que la distribución de los casos de ECV se encuentra en el rango de 0.3 y 1.3, encontrándose más dispersos. Por el contrario, la varianza se distribuye en los casos de ECV en un rango comprendido entre 0 y 0.14, siendo el rango de los casos saludables menor entre 0 y 0.04, aunque existe gran cantidad de casos con lo que se traslapan en dichas zonas.

En la Figura 6.21, la magnitud de los valores en magnitud de sesgo se tienden a agrupar en un intervalo entre -2 y 2, mostrando nuevamente los casos de ECV una dispersión mayor en el eje, mientras que para el caso de la Curtosis ilustrado en la Figura 6.20 los valores para casos sanos muestran valores inferiores a 2, tomando los casos de ECV de la misma forma una distribución mas dispersa a lo largo de dicha variable.

Comparativa Grupo sano vs ECV

Para concretar el panorama general de la acumulación de la muestra en conjunto de pacientes sanos vs pacientes patológicos, se exploran los datos de dicha muestra mediante un histograma y su respectivo ajuste a una distribución. Esta visualización en la figura 6.22 permite observar las tendencias de la distribución de frecuencia de los momentos obtenidos del intervalo R-R para los grupo de estudio.

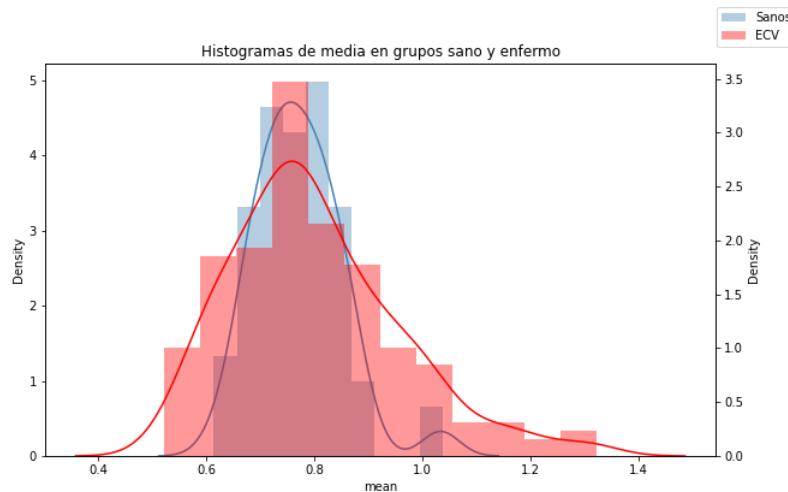


Figura 6.22: Histograma de la media en grupo sano y con ECV

De modo que, al obtener todas las distribuciones para un momento específico se desea conocer si tienden a agruparse a manera de aproximación. En segundo lugar, de esta observación se elige un caso que represente mejor al régimen Gaussiano del momento, y así continuar con el análisis de la evolución temporal de los momentos. En el Apéndice 10 se encuentran concentradas todas las visualizaciones resultantes de este proceso.

Media		
Característica	Sanos	ECV
Máximo	1.037998	1.320596
Mínimo	0.613441	0.522305
Media	0.770460	0.800874
DE	0.079624	0.157685

Cuadro 6.2: Características de la media

Varianza		
Característica	Sanos	ECV
Máximo	0.068124	0.141730
Mínimo	0.003351	0.000029
Media	0.019373	0.025145
DE	0.010455	0.030466

Cuadro 6.3: Características de la varianza

Sesgo		
Característica	Sanos	ECV
Máximo	1.182621	6.733672
Mínimo	-1.565596	-7.666674
Media	0.073459	1.783606
DE	0.613135	0.030466

Cuadro 6.4: Características de sesgo

Curtosis		
Característica	Sanos	ECV
Máximo	7.754840	122.099634
Mínimo	-0.945995	-1.755250
Media	0.965784	7.238784
DE	1.788370	15.698643

Cuadro 6.5: Características de la Curtosis

Habiendo generado las distribuciones de cada uno de los momentos en los respectivos grupos de estudio se realizó un análisis comparativo a partir de la descripción estadística de ellos y su representación en un espacio. Las características destacadas para determinar una diferenciabilidad inicial fueron el Máximo, Mínimo, Media, Desviación Estándar (DE).

Posteriormente, para determinar la similitud entre las distribuciones de probabilidad de los datos se realiza una prueba estadística de similitud o Kolmogórov-Smirnov. Este test no paramétrico mide la igualdad entre distribuciones continuas unidimensionales comparando su valor P . Una vez realizado este test, los resultados indican que la similitud entre las señales del mismo grupo es hasta de 65 % una comparativa con los grupos de distribuciones pertenecientes a las distintas patologías nos muestra resultados similares con un 63.23 % de similitud.

Análisis No Lineal

Para la segunda etapa de análisis, el objetivo fue analizar con técnicas no lineales las señales patológicas recopiladas de MIMIC-III en la base de datos descargada para las 3 patologías (FA, IAM y CC). Se aplicaron las técnicas de entropía aproximada, entropía muestral, DFA, exponente de Hurst, dimensión fractal de Higuchi y espectro de potencia a las ventanas obtenidas.

Ya que se trabajaría con señales RR de grandes longitudes y muchas en número se optó por adaptar multi-procesamiento, haciendo uso de todos los recursos del sistema para acelerar el resultado, el procedimiento se puede visualizar en la Figura 6.38.

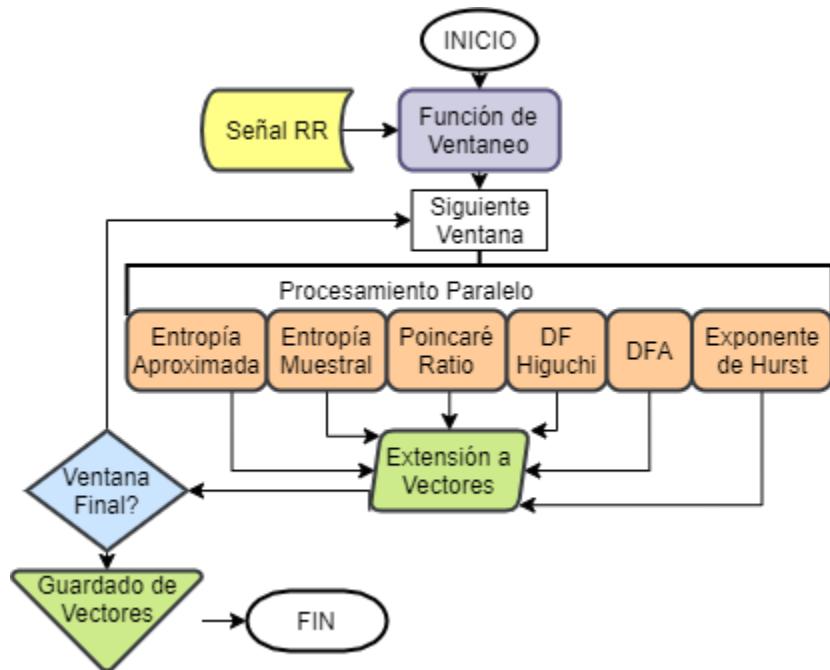


Figura 6.23: Diagrama de flujo del análisis no lineal

Análisis sobre fractalidad y no-linealidad de las señales

De manera similar al análisis realizado en la etapa anterior, se continúa con la idea de observar las gráficas resultantes del deslizamiento de la ventana diseñada sobre la señal de interlatido cardiaco, ahora extrayendo la evolución temporal de las métricas no-lineales.

A continuación se presentan en las figuras 6.24, 6.25, 6.26 y 6.27 presentan el resultado del ventaneo para todas las señales de la base DUMMY.

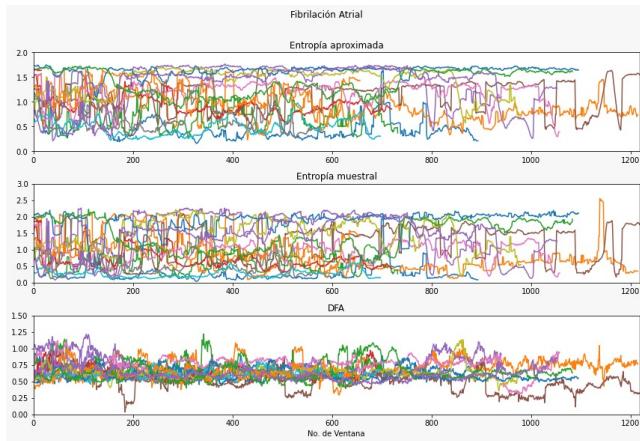


Figura 6.24: Gráfica de métricas no-lineales de FA

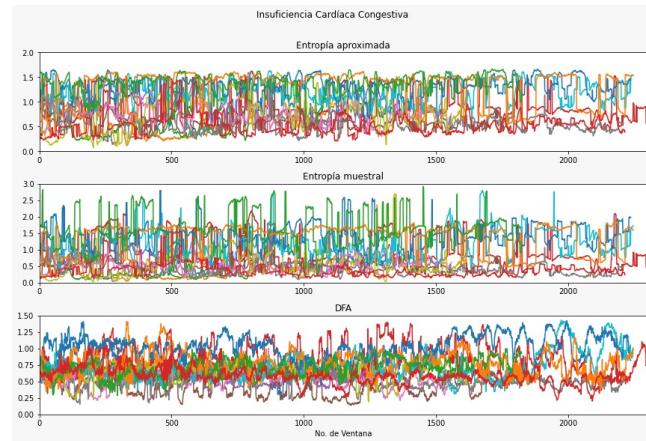


Figura 6.25: Gráfica de métricas no-lineales de ICC

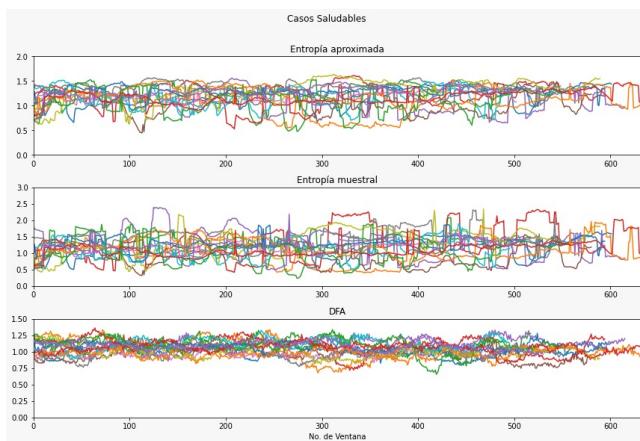


Figura 6.26: Gráfica de métricas no-lineales casos Saludables

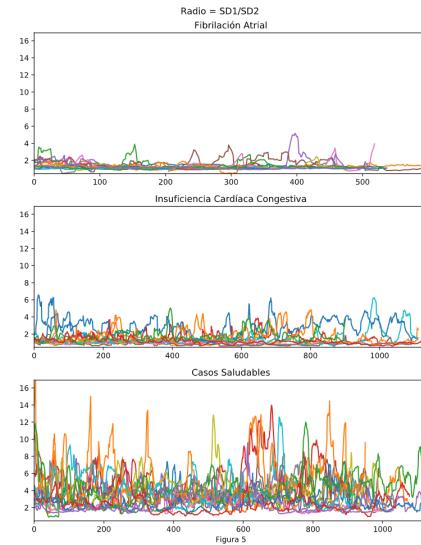


Figura 6.27: Gráfica de cociente R

Análisis de Poincaré

En la figura 6.28 se muestra el resultado de la construcción del mapa de retorno o gráfico de Poincaré de un registro de R-R saludable para la misma ventana de tiempo definida para el resto de las métricas.

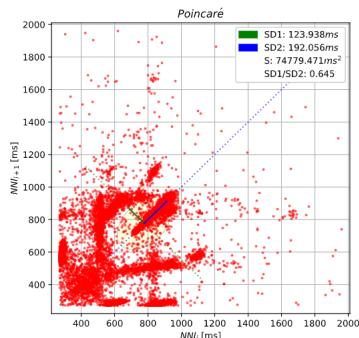


Figura 6.28: Gráfico de Poincaré para una señal saludable

Debido a la inconveniencia computacional de generar un gráfico de retorno tanto para DUMMY como para una base de mayor tamaño como es MIMIC-III se almacena de este análisis para las patologías y casos saludables. En la figura 6.27 se observa un gráfico del comportamiento del cociente R, obtenido del análisis del gráfico de Poincaré.

Análisis por Transformada Wavelet

Para el análisis por transformada wavelet se lleva a cabo para encontrar información de frecuencia real dentro de la señal de HRV. Para ello definimos una ventana de tamaño fijo que encierra la señal de HRV, restringiéndola en el tiempo y realizando la transformación para ajustar una ondeleta con el fin de modelar el comportamiento de HRV, sin embargo es importante notar que el modelado puede fallar, sobre todo en aquellas regiones en las que el comportamiento hace evidente que debido a la no periodicidad o no estacionariedad de la señal el ajuste a un modelo matemático se hace una labor mucho más compleja, La figura 6.29 muestra el ajuste de una ondeleta a la señal HRV, donde es notorio que estos métodos requieren de una búsqueda exhaustiva de una ondeleta o combinación de ellas que permita el ajuste completo de la señal.

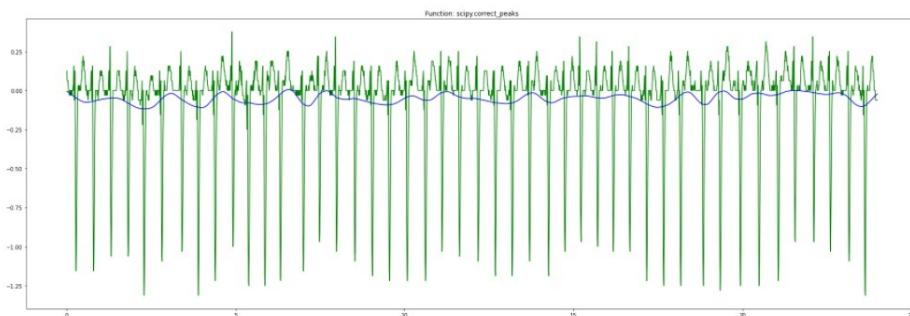


Figura 6.29: Transformada wavelet para registro sano de R-R

Inspección por medio del Espectro de Potencia de Wavelet

Dado que la aplicación del espectro de potencia wavelet (WPS) es muy similar a la densidad espectral de potencia tradicional basada en la transformada de Fourier, es comúnmente usada para encontrar eliminar componentes de alta frecuencia de la señal analítica que interfieran con la línea base y se ha comprobado su efectividad para detectar ruido de tipo $1/f$. Para definir un tamaño de ventana óptimo para cada señal analizada, se debe de inspeccionar el WPS, esto con la finalidad de encontrar intervalos de alta densidad de potencia los cuales se relacionan directamente con comportamiento fractal en la señal, por lo que se debe buscar un tamaño de ventana en un intervalo temporal que permita rescatar dichas zonas.

A continuación en las Figuras 6.32,6.30 y se muestra el espectro de potencia de wavelet para casos de FA e ICC, en dónde se puede notar una acumulación de estos eventos fractales denotados por las franjas más claras, sin embargo considerando la escala en el eje vertical que daría una ventana adecuada, es posible ver que incluso para los casos patológicos se tiene dichos eventos en una ventana cercana a los 600 datos.

Pero si se desea obtener la ventana óptima de cada registro, esta inspección es necesaria con sus debidos ajustes a la longitud del mismo.

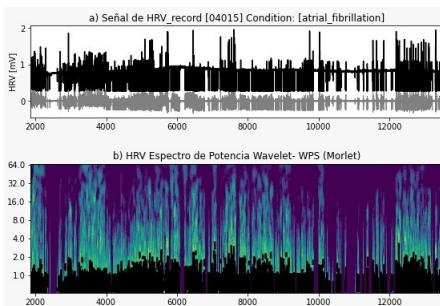


Figura 6.30: WPS de registro de AF

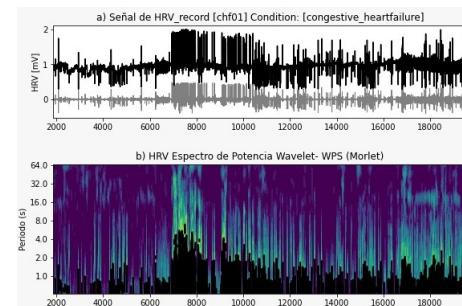


Figura 6.31: WPS de registro de ICC

De manera análoga puede observarse el comportamiento de la WPS para un registro saludable en la figura 6.31. Sin embargo, debido a la complejidad de la señal saludable, es mucho mayor la cantidad de ajustes necesarios para lograr visualizar estos eventos.

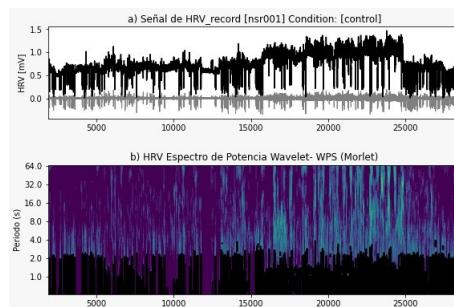


Figura 6.32: WPS de registro Sanos

Generación de archivos csv

De la misma forma que en la primera etapa, el ventaneo resultante de aplicar las técnicas no lineales en la ventana deslizante produce vectores de diferentes longitudes, que será almacenado en un archivo csv, sin embargo como se necesitan características puntuales de la señal se requirió extraer estas métricas por medio del cálculo de la media, varianza y sesgo del vector resultante para cada una de las técnicas.

Lectura y limpieza de datos

El csv generado es leído nuevamente por la librería pandas de python, convirtiéndose en un objeto dataframe para su posterior análisis. Mediante una etapa de limpieza de datos se descartaron valores incorrectos encontrados en el archivo csv, quedando un total de 29 columnas y 18222 filas. Tras esta limpieza la tabla en formato dataframe obtenida para cada una de las técnicas cuenta con el siguiente formato:

Técnica utilizada		
Media	Varianza	Sesgo
m1	v1	s1
m2	v2	s2
m3	v3	s3

Cuadro 6.6: Estructura de la tabla obtenida de la recopilación de las técnicas no lineales

Gráficos de dispersión

En las Figuras 6.33, 6.34, 6.35 y 6.36, se muestran los gráficos de dispersión resultantes de comparar las técnicas no lineales aplicadas en esta etapa, el resto de gráficos pueden consultarse en el apéndice.

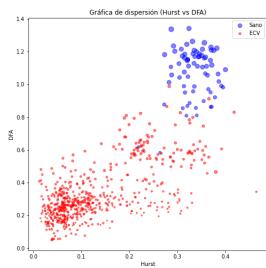


Figura 6.33: Gráfico de dispersión del exponente de Hurst vs DFA.

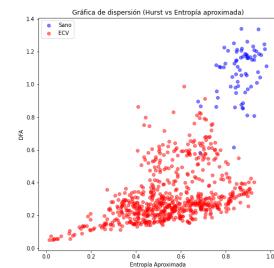


Figura 6.34: Gráfico de dispersión DFA vs entropía aproximada.

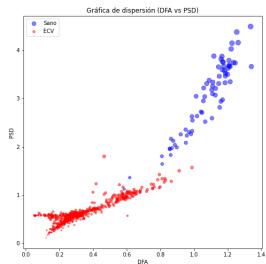


Figura 6.35: Gráfico de dispersión DFA vs PSD.

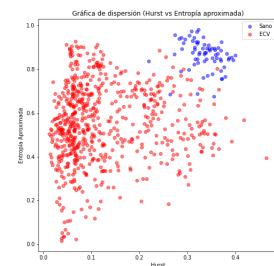


Figura 6.36: Gráfico de dispersión Hurst vs entropía aproximada.

En las Figura 6.33 se aprecian dos cúmulos principales, siendo el de color rojo los registros para registros diagnósticados con ECV y los de color azul para la base de datos de pacientes saludables. Los pacientes con ECV cuentan con valores en magnitud inferiores de dimensión fractal y de exponente de Hurst, comparado con los registros saludables que muestran una clara tendencia en magnitud superior para ambos análisis localizándose en intervalos para Hurst mayores a 0.25 y de dimensión fractal superior a 0.8

Para el caso de la entropía aproximada y PSD ilustrados en las Figuras 6.34 y 6.36 los registros saludables muestran valores de magnitud de entropía superiores a 0.8 y para la técnica de PSD un valor superior a 2. Estas gráficas hacen evidente primero, que los valores de los registros sanos tienden a tener valores mayores en magnitud para cada una de las técnicas aplicadas, y segundo que se encuentra una correlación positiva para al comparar los resultados del exponente de Hurst, DFA y PSD.

Análisis no lineal (Grupo sano vs ECV)

Como segundo acercamiento para encontrar una tendencia de separabilidad en una primera exploración con los grupos de pacientes sanos y de ECV, se introduce el uso de técnicas no lineales y de sistemas complejos, en concreto, el cálculo de la dimensión fractal por medio de técnicas como DFA, Higuchi, entropía de información, utilizando entropía aproximada y muestral, así como el uso del exponente de Hurst para medir la aleatoriedad de las señales y por último el espectro de potencia de las mismas.

De modo que, al aplicarse a las ventanas deslizantes obtenidas sean capaces de mostrar criterios de separabilidad superiores a los vistos anteriormente con el uso de los 4 momentos estadísticos. A continuación en la Figura 6.37 se presenta el histograma para ambos grupos de estudio utilizando como base el cálculo de dimensión fractal por DFA, el resto de histogramas para el resto de técnicas se muestran en el apéndice .

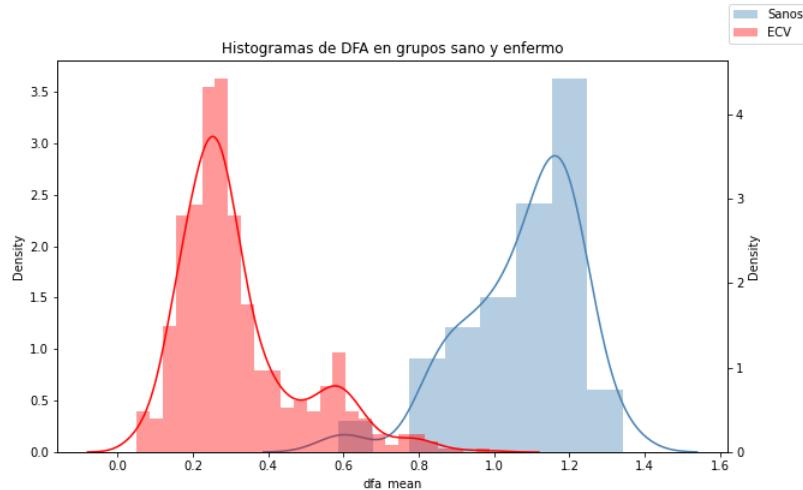


Figura 6.37: Histogramas de la dimensión fractal calculada por DFA en grupo sano y con ECV

Clasificación

Reducción dimensional por PCA

Empleando los archivos csv generados en pasos anteriores se extraen a un dataframe por categoría, donde se proveen 22 características al análisis, como resultado de las 6 técnicas no lineales aplicadas. Configurando el número variable N como el número de componentes esperados a la salida, se procesan secuencialmente los datos para cada patología.

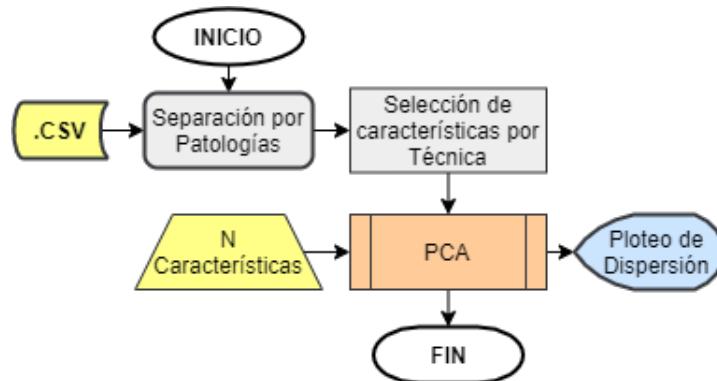


Figura 6.38: Uso de PCA a partir de CSV

Modelo de clasificación supervisado

Para verificar la eficiencia de estas técnicas para clasificar entre estado saludable y patológico, se implementó un algoritmo clasificador de aprendizaje supervisado, el cual se entrena con la información recopilada sobre los momentos estadísticos (media, varianza y sesgo) en la señal resultante de la aplicación de técnicas no lineales sobre la señal HRV.

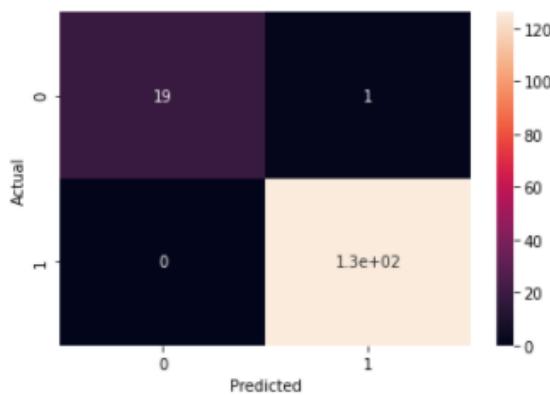


Figura 6.39: Matriz de confusión obtenida del proceso de prueba del algoritmo de clasificación entrenado con un tamaño de prueba de $n = 150$

El algoritmo que decidió implementarse fue Extra Trees clasifier, siendo un modelo de clasificación basado en árboles de decisión cuya entrada fueron las 22 características recopiladas de la señal, la matriz de confusión obtenida al realizar el proceso de pruebas se muestra a continuación en la Figura 6.39.

Adicionalmente se realizaron las curvas ROC y r-c las cuales son herramientas útiles para predecir la probabilidad de un resultado binario.

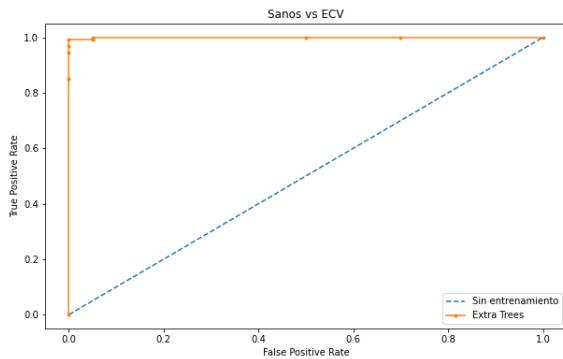


Figura 6.40: Curva ROC del algoritmo de clasificación implementado para diferenciar sanos vs ECV.

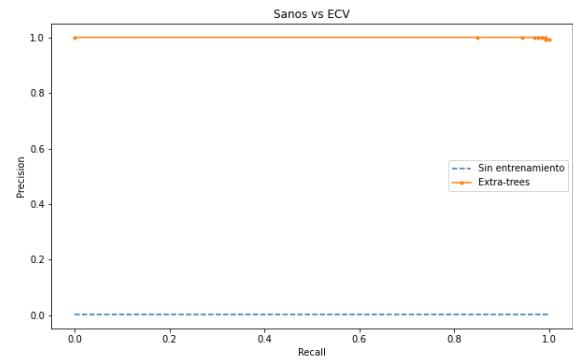


Figura 6.41: Curva R-C del algoritmo de clasificación implementado para diferenciar sanos vs ECV.

Análisis no lineal para patologías cardíacas (FA, CC e IM)

Habiendo demostrado la efectividad de las técnicas no lineales en brindar criterios que puedan ser utilizados para la clasificación de grupos saludables y con ECV, el objetivo siguiente del trabajo se centra en analizar registros de FA,CC e IM para evaluar la efectividad de estas técnicas en diferenciar entre estados patológicos, buscando obtener criterios de diferenciación que puedan ser identificados como precursores al evaluar señales electrocardiografías y mostrar su desplazamiento hacia alguna agrupación .

En la figura 6.42 se muestran los histogramas de la media de la dimensión fractal calculada por DFA junto con la función de densidad para los 3 grupos de estudios, como se puede apreciar, estos histogramas muestran que no existe un criterio suficiente para su clasificación por medio de este análisis, aunque el caso para FA muestra una media desplazada hacia la derecha y un cambio de densidad en un rango de valores entre 0.4 y 0.6

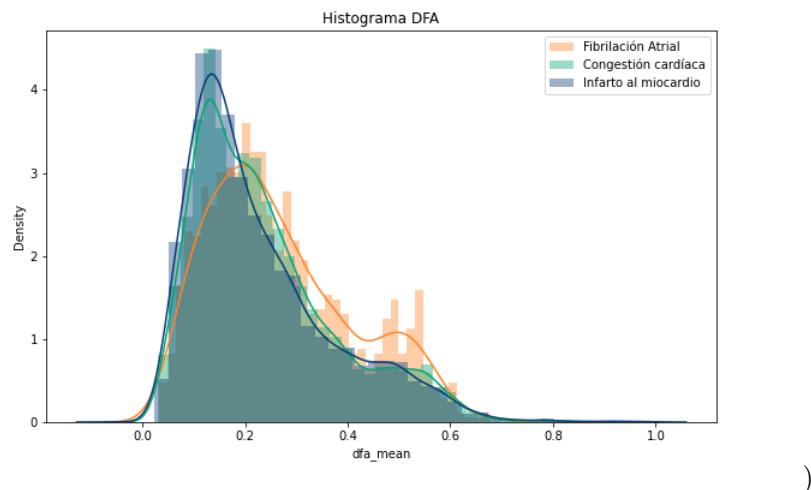


Figura 6.42: Histograma mostrando la técnica de dimensión fractal por DFA para las 3 patologías

Para el caso de entropía muestral, ilustrado en la figura 6.43, estas características cuentan de nuevo con media similar, diferenciándose en conjunto por el valor de la densidad de probabilidad en su punto central, siendo el caso para FA de nuevo diferente respecto a CC e IM, al presentar un cambio de densidad más evidente en el rango comprendido entre los valores de 0.4 y 0.6. El resto de histogramas se muestran en el apéndice haciendo notorio que por el uso de una sola técnica aplicada no se encuentra una separación significativa de los datos.

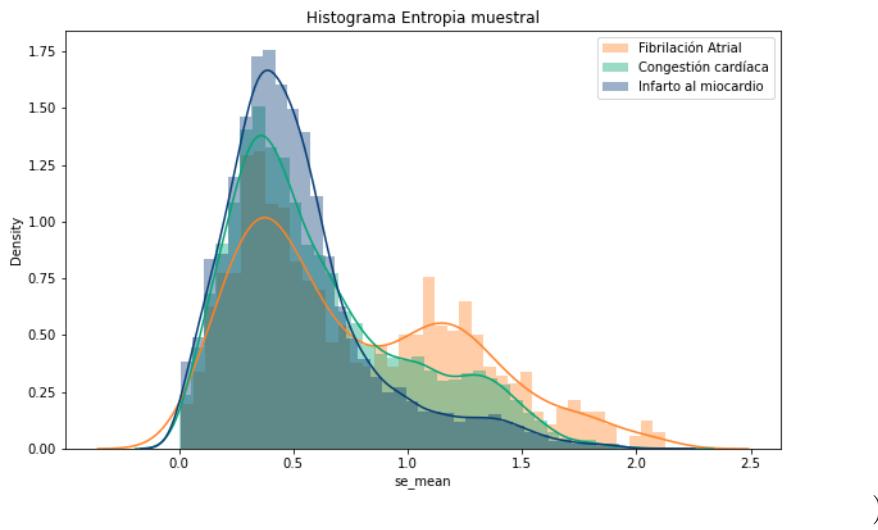


Figura 6.43: Histograma mostrando la técnica de entropía muestral para las 3 patologías)

Gráficas de dispersión

En las Figuras 6.44, 6.45, 6.46 y 6.47 se muestran las gráficas de dispersión resultantes al comparar entre algunas de las distintas técnicas no lineales aplicadas en esta etapa, el resto de gráficos puede consultarse en el apéndice.

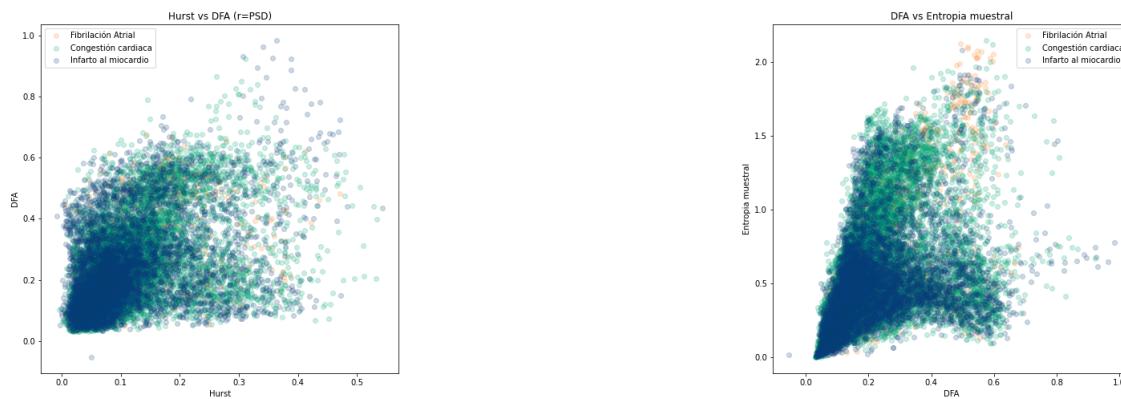


Figura 6.44: Gráfico de dispersión del exponente de Hurst vs el índice fractal por DFA.

Figura 6.45: Gráfico de dispersión del índice fractal por DFA vs entropía muestral.

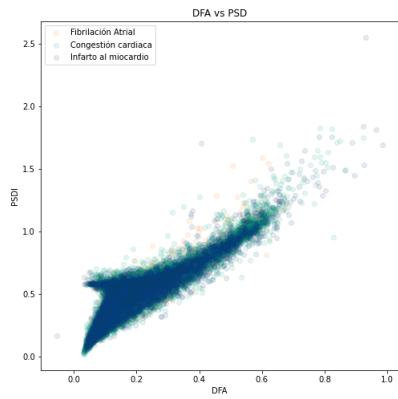


Figura 6.46: Gráfico de dispersión del índice fractal por DFA vs Espectro de Potencia.

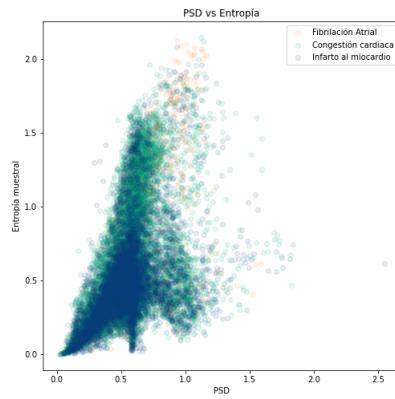


Figura 6.47: Gráfico de dispersión del espectro de potencia vs entropía muestral.

Reducción dimensional por PCA Entre patologías

Al no encontrarse separación significativa al comparar resultados entre 2 técnicas en los grupos de estudio se realizó una reducción dimensional por PCA, esto con el fin de reducir el número de características empleadas (22) a solo 3 componentes principales y encontrar si entre estas existían diferencias. En la figura 6.48 se muestra la gráfica de dispersión obtenida comparando las 3 componentes principales para los 3 grupos.

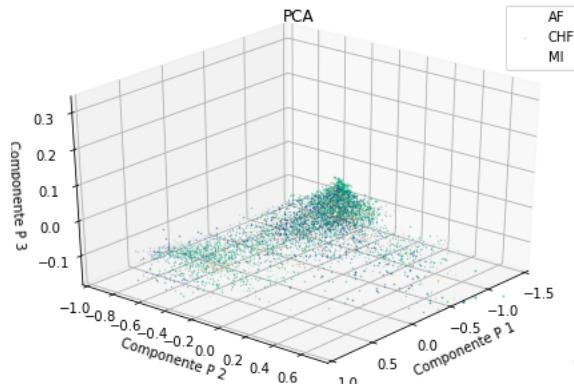


Figura 6.48: Gráfica de dispersión para las 3 componentes principales obtenidas por el PCA

Esto hace evidente que no es posible diferenciar las 3 patologías con el uso de 3 componentes principales, dado el total de 22 características iniciales se requiere la integración de un número mayor de componentes que permita describir de mejor forma los fenómenos, para lo cual se necesita del algoritmo de clasificación supervisado con el fin de evaluar el desempeño de estos componentes superiores para la clasificación de FA, IM e ICC.

Clasificación supervisada

Utilizando el mismo algoritmo de clasificación supervisado “Extra trees classifier” se entreno el modelo para distinguir las 3 patologías a partir de la reducción dimensional dada por la técnica de PCA. El desempeño de este algoritmo clasificador en la etapa de prueba con un 20 % de los registros totales, utilizando un total de 9 componentes mediante la reducción dimensional por PCA, se obtuvo una efectividad del 90 % tal como se muestra en la matriz de confusión de la Figura 6.49.

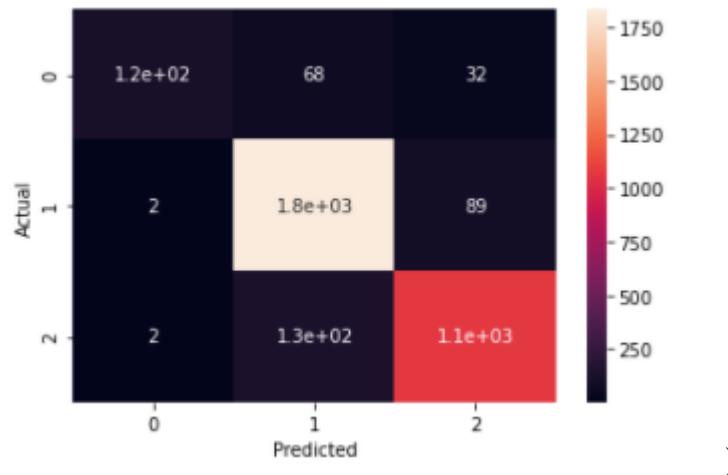


Figura 6.49: Matriz de confusión resultante del modelo de clasificación para las 3 patologías con una muestra de entrenamiento del 20 %

Se encontró a su vez que existía una correlación entre el número de componentes del PCA y la efectividad del modelo para la clasificación. En el cuadro 6.7 se muestran algunas de las eficiencias encontradas para diferentes números de componentes analizados, mientras en la Figura 6.50, se ilustra el comportamiento de esta relación.

Componentes principales	Eficiencia (%)
1	48
2	54
3	58
4	66
5	79
6	81
7	87
8	88
9	90
10	90.2
11	90

Cuadro 6.7: Componentes principales analizadas vs eficiencia del modelo

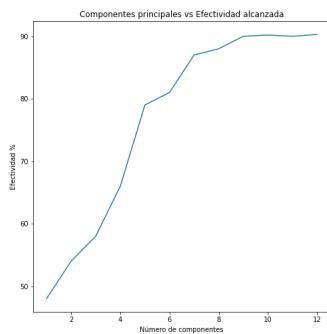


Figura 6.50: Relación entre el número de componentes principales analizadas y la eficiencia del modelo

Comparación entre FA e ICC

Para validar el procedimiento hasta ahora realizado se decidió evaluar el desempeño de la clasificación comparando patologías individualmente, en este caso el algoritmo de clasificación fue únicamente entrenado con datos de registros para FA e ICC, con una reducción dimensional de 9 componentes, se obtuvo la siguiente matriz de confusión de la etapa de prueba ilustrada en la Figura 6.51, obteniendo una eficiencia del 95 %.

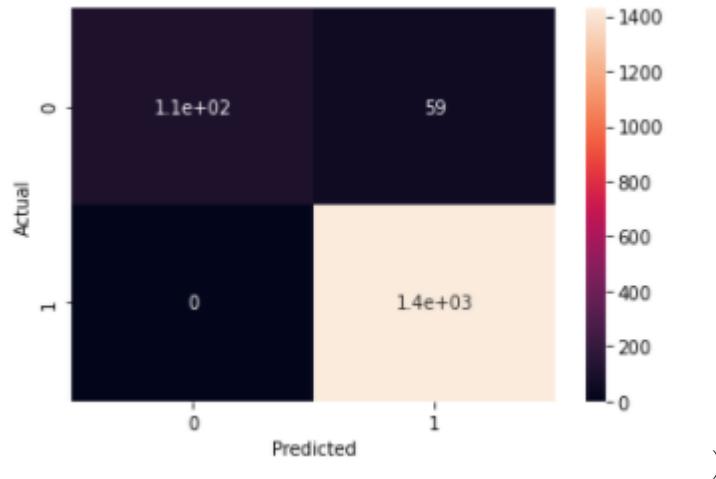


Figura 6.51: Matriz de confusión resultante de la clasificación del algoritmo de FA e ICC

También se obtuvo la curva ROC y R-C para el modelo entrenado, tal como se muestra en las Figuras 6.52 y 6.53.

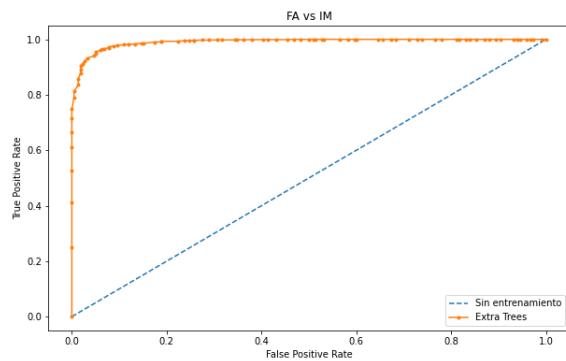


Figura 6.52: Curva ROC para el modelo de clasificación entrenado con los datos de registros de FA e ICC

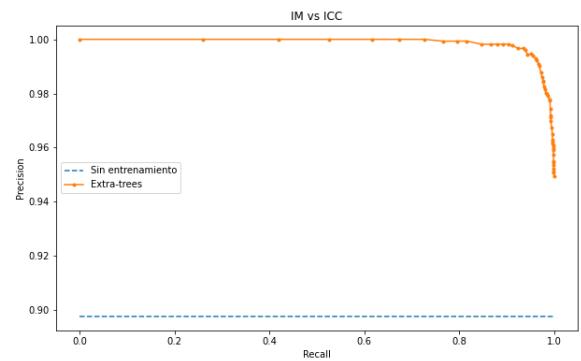


Figura 6.53: Curva R-C para el modelo de clasificación entrenado con los datos de registros de FA e ICC.

Comparación entre FA e IM

De la comparación efectuada entre FA e IM, se entreno al modelo siguiendo la misma lógica anterior, reportando la matriz de confusión para las pruebas realizadas en la Figura 6.54, obteniéndose una eficiencia del 96 %.

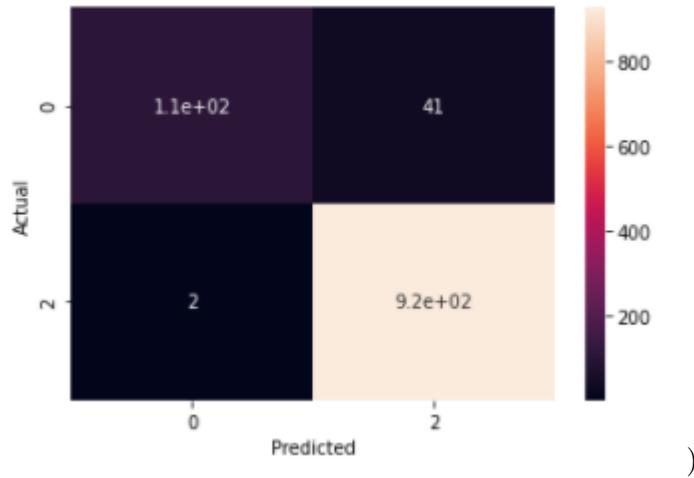


Figura 6.54: Matriz de confusión resultante de la clasificación del algoritmo de FA e IM

Las curvas ROC y R-C para el modelo se muestran en las Figuras 6.55 y ??.

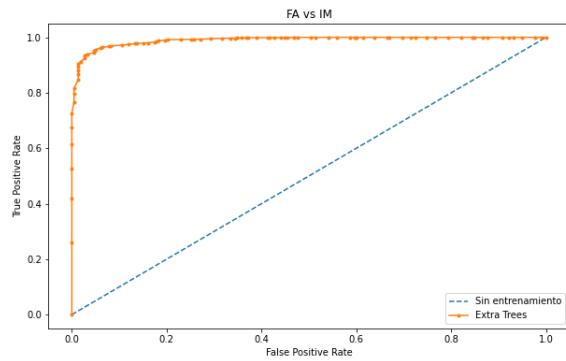


Figura 6.55: Curva ROC para el modelo de clasificación entrenado con los datos de registros de FA e IM

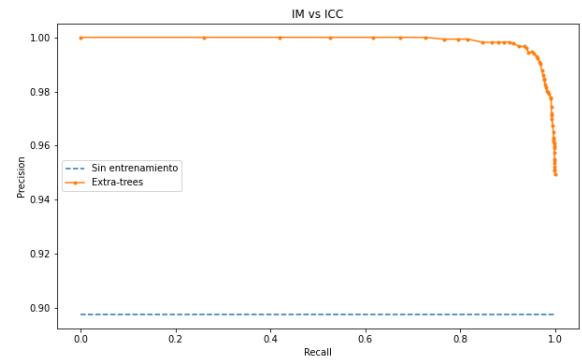


Figura 6.56: Curva R-C para el modelo de clasificación entrenado con los datos de registros de FA e IM.

Comparación entre ICC e IM

De la comparación efectuada entre ICC e IM, se entreno al modelo con ambos grupos, reportando la matriz de confusión para las pruebas realizadas en la Figura 6.57, obteniéndose una eficiencia del 91 %.

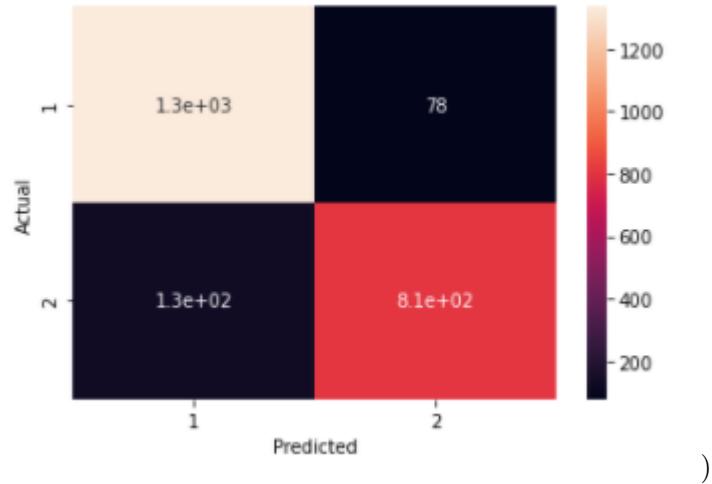


Figura 6.57: Matriz de confusión resultante de la clasificación del algoritmo de ICC e IM

Las curvas ROC y R-C para el modelo se muestran en las Figuras 6.58 y 6.59.

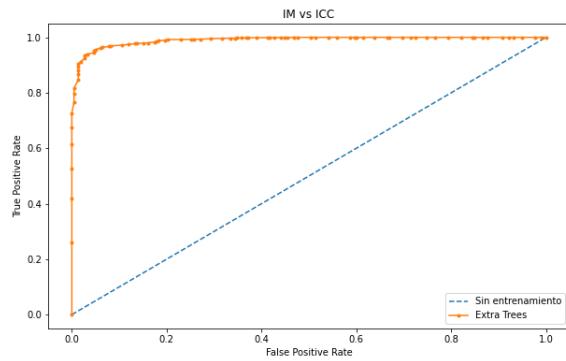


Figura 6.58: Curva ROC para el modelo de clasificación entrenado con los datos de registros de IM e ICC

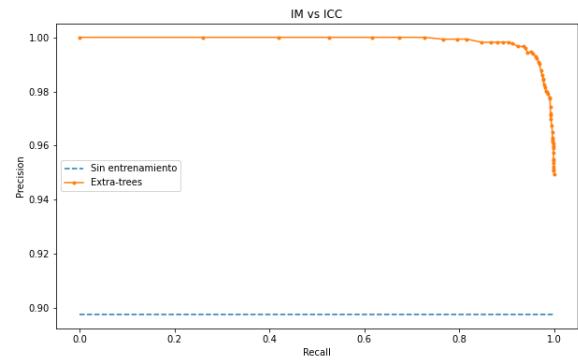


Figura 6.59: Curva R-C para el modelo de clasificación entrenado con los datos de registros de IM e ICC

De las comparaciones anteriores, el cuadro 6.8 compara cada uno de los porcentajes de eficiencia obtenidos en el modelo.

Patologías	Eficiencia (%)
FA vs ICC	95
FA vs IM	96
IM vs ICC	91

Cuadro 6.8: Porcentajes de eficiencia obtenidos

Capítulo 7

Análisis e Interpretación de Resultados

Análisis Lineal

Las técnicas aplicadas en este etapa mostraron no poder discriminar entre patologías tan eficientemente como en la etapa anterior para diferenciar entre pacientes sanos y con ECV al solo utilizar una característica como lo indican los histogramas reportados en las Figuras 6.42 y 6.43, aunque los histogramas de FA mostraban cambios en su función de densidad de probabilidad en cierto rango de valores respecto al resto.

Los gráficos de dispersión al comparar entre dos técnicas tampoco mostraron diferenciabilidad entre las mismas, por lo que dado el número de características obtenidas tras el proceso de ventaneo dinámico se optó por realizar una reducción dimensional que fuera capaz de comprimir dicha información y presentarla en un número de componentes menor, sin embargo al utilizar 3 componentes principales tampoco fue posible obtener un gráfico que mostrará las diferencias entre los 3 grupos de estudio.

Al inferir que sería posible que la diferenciación se encontrará al obtener un número de componentes mayores, se decidió entrenar al modelo de aprendizaje con distinto número de componentes como lo indica el cuadro 6.7, encontrándose una relación entre las componentes utilizadas y la eficiencia de clasificación, por lo que se sugiere que existen criterios de diferenciación entre los 3 conjuntos mas allá de las 3 dimensiones que es posible visualizar con gráficos.

De la misma forma, las comparaciones realizadas entre patologías y las eficiencias reportadas en el cuadro 6.8, además de las ligeras diferencias mostradas por los histogramas, se puede deducir que los registros de FA se distinguen de IM e ICC en mayor medida, al reportarse un incremento de eficiencia del 5 % en su efectividad al clasificarse contra el resto.

Análisis no Lineal

En la Figura 6.37 y en el resto de histogramas referenciados en el apéndice, se hace notoria la separabilidad de las gráficas para los conjuntos de técnicas en pacientes sanos y con ECV por medio de un análisis no lineal, teniendo los casos saludables una media establecida de 1.2 para el análisis por DFA, mientras los casos patológicos una media en 0.31 siendo en una primera instancia la pérdida de fractalidad en la señal R-R un factor con tendencia a patología cardíaca encontrada en los grupos de estudio.

En las gráficas de dispersión 6.33,6.34,6.35 y 6.36, también se muestran tendencias de agrupación para ambos grupos de estudio, en cuanto a la entropía aproximada, los casos con diagnóstico saludable muestran valores mayores en magnitud superiores a 0.8, mientras la mayoría de los registros para pacientes con ECV se encuentran por debajo de este valor. En cuanto al espectro de potencias, el grupo de diagnóstico saludable se encuentra en valores superiores a 1.2, siendo el grupo de ECV el que presenta menos energía en su señal.

En la prueba del algoritmo clasificador supervisado (“Extra Trees Clasifier”), se obtuvo con un tamaño de prueba 150 registros escogidos aleatoriamente, una efectividad del 99.31 %, siendo este valor reportado óptimo para diferenciar entre los dos estados de esta etapa.

Los resultados obtenidos son concluyentes en que es posible identificar con alta efectividad registros

de ECG provenientes de pacientes saludables y con diagnóstico de ECV por medio de este conjunto de técnicas no lineales y con la implementación de un algoritmo de aprendizaje supervisado, mismos que se han reportado con anterioridad en la literatura. En general las características aportadas por las técnicas de DFA, exponente de Hurst, Higuchi, PSD, entropía muestral y aproximada muestran tener una tendencia en señalar la aproximación de registros de prueba a uno de los dos estados, sentando las bases para explorar su utilidad en la diferenciación de patologías y su uso como precursores asociados a ECV.

Modelo de clasificación

El primer modelo de aprendizaje supervisado fue utilizado para evaluar el desempeño de las métricas no lineales en la diferenciación de estado saludable y patológico. La curva ROC ilustrada en la figura 6.40 muestra que el desempeño del modelo es óptimo al compararse con uno aleatorio, esto hace posible que se alcancen precisiones altas al incrementarse el ratio de positividad a una tasa mayor que el ratio de negatividad. La curva p-r de la Figura 6.41 muestra de nuevo la comparación del modelo con uno aleatorio, esta vez graficándose la precisión alcanzada contra las llamadas incorrectas realizadas, concluyendo el modelo supera significativamente a uno utilizado con características aleatorias.

En los modelos consecuentes implementados para las 3 patologías de estudio (FA,ICC e IM) se obtuvo una precisión del 90 % al comparar entre las 3 clases, mientras que para los cruces de patologías, FA se logró diferenciar del resto con una efectividad de hasta el 95 %. Como se muestra en las Figuras 6.52, 6.53, 6.55 y ?? las curvas ROC y r-p muestran desempeños superiores al entrenamiento con características aleatorias para diferenciar FA.

Las 9 componentes principales utilizadas en la etapa de entrenamiento junto con los resultados de eficiencia en la etapa de clasificación demuestran que existen criterios de separabilidad para estas 3 patologías que brindan los valores en magnitud de los gráficos de Poincaré, la entropía muestral, entropía aproximada,dimensión fractal por DFA, Higuchi y exponente de Hurst, aunque dado el número de dimensiones analizado no fue posible visualizar gráficos contenidos en hiperplanos. Ya que las características se derivan del análisis mediante sistemas complejos, estas han resultado ser útiles para modelar transiciones entre estados en sistemas, por lo que como resultado pre-eliminar se ha logrado caracterizar precursores que puedan implementarse en modelos de evaluación que apoyándose en los valores de magnitudes definidos en el estado crítico, sean capaces de calcular un grado de similitud para el estado patológico y posteriormente entre FA, ICC e IM

Capítulo 8

Discusión

En el presente estudio se ha desarrollado una metodología que integra algoritmos de procesamiento, de análisis lineal y no lineal, y de machine learning, para la caracterización y clasificación de secuencias de interlatido cardiaco provenientes de individuos con patologías cardíacas y sujetos sanos. La base de datos MIMIC-III, la cual sido objeto de diversas investigaciones, sentó las bases para la realización de este trabajo, al contar con un acervo de datos clínicos que asciende a mas de 60mil pacientes.

El contar con un catálogo robusto de este tipo de datos, se vuelve un reto, sobretodo en el caso de países en desarrollo como México. Debido a la naturaleza de los datos, la gran mayoría de estos son de difícil acceso para investigadores, en su mayoría teniendo que seguir protocolos de acceso rigurosos y con demoras en su aprobación, sin embargo, esfuerzos como este han logrado conciliar dentro de una normativa bioética un protocolo de acceso mas eficiente a partir del cual se han derivado múltiples investigaciones.

El análisis de señales cardíacas se beneficia notablemente de estas acciones, ya que es posible extraer métricas de mayores muestras poblacionales y se cuenta con registros suficientes para validar hipótesis planteadas con anterioridad.

En este panorama, se abordo el problema de la caracterización de precursores para enfermedades cardiovasculares (FA, ICC e IM) en señales R-R , buscando incorporar el paradigma de la complejidad y el uso de algunas técnicas dentro de su margen como lo son el análisis por fractales, entropía y transformada de ondeletas. Tomando en consideraciones resultados reportados con anterioridad en su eficiencia para diferenciar alteraciones y al ser un problema que no podía ser abordado por técnicas convencionales de tipo lineal como los momentos estadísticos.

Habiendo recopilado 18,222 registros para las 3 patologías estudiadas provenientes de MIMIC-III, existió un gran costo computacional asociado para su procesamiento, el cual no hubiera podido ser posible de no ser porque se contó con el acceso al servidor del laboratorio de Sistemas Complejos de la UPIITA, mostrando que este tipo de estudios consumen una gran cantidad de recursos que un equipo de computo individual convencional no podría realizar en el mismo tiempo debido a sus características.

Otro reto asociado a la gran cantidad de señales de estudio fue el obtener un tamaño de ventana óptimo con el cual proseguir con los distintos análisis, ya que para encontrar características como fractalidad, se sugiere visualizar el espectro de potencias de ondeletas de un registro para ajustar el tamaño de ventana de acuerdo a los intervalos con mayor densidad de potencia. Esto hace una tarea complicada para visualizar cada registro de manera individual, se optó por elegir un tamaño de ventana de 500 datos, ya que en este caso, las señales exhibieron índices fractales estables, y consistentes con los valores reportados en la literatura.

Al aplicar este conjunto de técnicas por medio de ventaneo dinámico de señales, si bien se logró aplicar en todos los registros, una problemática asociada fue el comprimir la información de la señal resultante de la ventana a datos puntuales que conservaran las características y pudieran ser ingresados a un algoritmo de ML, por lo que se decidió utilizar momentos estadísticos para caracterizarlas, sin embargo, el uso de otras técnicas como la reducción dimensional deben explorarse.

A partir de los resultados obtenidos se pueden trazar conclusiones que involucren, en el primero de los análisis, la diferenciación entre estado patológico y saludable de las señales a partir de aplicar técnicas no lineales y de sistemas en complejos, demostrando que la pérdida de dimensión fractal en una señal puede interpretarse como un precursor para ECV, teniendo mas trabajos reportados en la literatura que concluyen una relación directa.

Para las 3 patologías estudiadas (FA, ICC e IM) el panorama resulta ser mas complicado, las carac-

terísticas puntuales extraídas del ventaneo de la señal no muestran en una primera instancia tener criterios de separabilidad unidimensionales o bidimensionales, al aplicarse reducción dimensional por PCA fue posible con 9 componentes principales llegar a una precisión de clasificación por ML del 90 %, indicando que estas características aportan una gran carga para caracterizar una señal y que, el abordar esta problemática mediante el paradigma de la complejidad puede ser un camino viable de investigación con mayores registros recopilados.

Tomando en cuenta las limitaciones actuales para el registro de alta definición de millares de señales y su recopilación se sabe que a futuro, estos recursos se desarrollarán más en el mundo gracias a el aporte en el campo médico y de salud que se obtiene de su análisis, por lo que como investigaciones exploratorias como éste suman importancia a dicha tarea, aún cuando queda mucho por mejorar desde la detección algorítmica de latidos en señales ECG, donde aún existen malas detecciones por la variabilidad de registros, hasta la aplicación de técnicas para extracción de características, donde cuestiones como el tamaño de ventaneo y la aplicación ordenada de modelos pueden generar resultados aún por explorar.

Capítulo 9

Conclusiones

Nuestros resultados, han mostrado la utilidad de combinar los llamados análisis lineales y no lineales con los algoritmos de ML, para la caracterización de secuencias complejas como las series de interlatido cardiaco. Específicamente, hemos sentado evidencia de la funcionalidad de estos métodos para la comparación entre señales normales y patológicas, dando como resultado la posibilidad de establecer criterios de diferenciación, que al ser entrenados en un modelo de aprendizaje supervisado, se obtengan eficiencias de hasta el 99 % para su clasificación. Siendo lo anterior parte fundamental del objetivo del proyecto, y con ello sentar las bases de su utilidad en la caracterización de patologías.

Las técnicas mencionadas anteriormente en su conjunto proveen también criterios de diferenciación para las patologías de FA, ICC e IM, al reportar resultados en el algoritmo de aprendizaje supervisado del 90 %. Con esto podemos establecer la relación de sus valores en magnitud como posibles precursores para las patologías de interés, dado que estas técnicas permiten mediar la transición entre estados en una señal. Se cuenta con parámetros que pueden ser utilizados por un sistema para calcular su aproximación con los estados de unidades de cuidados intensivos.

A pesar de tener resultados adecuados con la técnica de reducción dimensional por PCA, este método no es el único que puede emplearse para realizar este tipo de reducciones, su ventaja recae en su fácil implementación. Al aplicar un método de reducción mucho mas sofisticado, se abre la posibilidad del uso de una menor cantidad de dimensiones para la correcta visualización de las características en un gráfico 3D.

La técnica de transformada de ondeletas, en un inicio planteada dentro de las técnicas a utilizar para clasificación, se descartó, dado que la aplicación ideal encontrada para ésta se centra mayormente en la fase de detección de picos para la obtención de la señal R-R y no se observaron propiedades significativas que se pudieran extraer de su análisis directo.

El tamaño de la ventana deslizante para las señales se estableció en un valor de 500 datos de lectura, este valor se consideró al realizar pruebas con señales controladas aleatorias las cuales arrojaron resultados esperados para el cálculo del exponente de Hurst y la dimensión fractal, sin embargo, se vislumbra la posibilidad de explorar diferentes tamaños de ventana de acorde a la duración total del registro y la patología, dado que aún no existen criterios establecidos para realizar estos ajustes.

Dado que se utilizó un modelo de aprendizaje supervisado para la clasificación, es posible de igual forma, implementar modelos no supervisados, los cuales sean capaces de agrupar los registros sin tener valores etiquetados y poder alcanzar mejoras en la eficiencia de clasificación.

Aunque se analizaron en total una cantidad de 18,222 registros de ECG, mayor a cualquier análisis reportado en el estado del arte para las patologías de FA,ICC e IM, se considera los resultados obtenidos son por ahora pre-eliminarios y deben de acotarse dentro del margen del grupo de estudio. Se plantea que su utilidad debe seguirse investigando dentro de bases de datos de investigación clínica.

En este trabajo se propusieron metodologías para la obtención de precursores entre señales patológicas y no patológicas, logrando la clasificación en un 99 % y cumpliendo el propósito principal, dando pie a mejorar la clasificación entre patologías en futuros estudios con datos más robustos.

Capítulo 10

Recomendaciones y/o Sugerencias para investigaciones futuras

El desarrollo de este trabajo demostró la necesidad de un procesamiento computacional elevado, el cual fue la mayor limitante para explorar diversas formas de aplicar las técnicas presentadas en el trabajo, pues el tiempo requerido por los equipos de cómputo para finalizar su ejecución excedía el límite establecido para la fecha de entrega.

Es por ello que la actual investigación contiene aún un amplio abanico de posibilidades, temas por explorar, y áreas de oportunidad para su mejora. A continuación se discuten algunas de estas áreas, así como algunas particularidades que podrían enriquecer a este trabajo en investigaciones futuras.

El uso de el espectro de potencia de wavelet para determinar el tamaño de ventana de observación en el cual se hace evidente el comportamiento fractal de la señal representa un trabajo de automatización en sí mismo. Debido a que realizar dicha inspección visual cada uno de los registros resulta una tarea gigantesca cuando se trata con bases de datos robustas; es por ello que algún método de automatización que logre detectar por cada registro en la base de datos un umbral óptimo en el espectro de potencia de wavelet en dónde es posible determinar un tamaño de ventana óptimo para el análisis de dicha señal y lo asigne como parámetro de entrada al tamaño de una función de ventaneo Dinámico representaría un avance con respecto a los métodos de ventaneo reportados en esta investigación.

Otra área de oportunidad para la investigación futura es la exploración de algoritmos no supervisados mas robustos existentes en la actualidad, como aquellos que pertenecen al marco del aprendizaje profundo (Deep learning), algoritmos genéticos, así como redes neuronales de tipo Autoencoder como lo son los LSTM y Transformers reportan en la literatura actual resultados interesantes que pueden aportar a esta investigación. Con la consideración de que estos algoritmos suelen ser más costosos computacionalmente hablando y requieren de equipos mucho mas robustos y avanzados para trabajar de manera óptima. Otros avances notables se encuentran en la reducción dimensional, actualmente existen técnicas basadas en conceptos mucho más avanzados como topología de datos para hacer el proceso de reducción dimensional mucho más eficiente. La exploración de algoritmos como Uniform Manifold Approximation and Projection (UMAP) podría brindar una oportunidad de mejoría en este aspecto con miras al aumento de la eficiencia de la clasificación reportada en esta investigación.

Bibliografía

- [1] S. Israel, A.Cheng and J. Irvine, “Ecg to identify individuals,” *Pattern Recognition*, 2005.
- [2] World Health Organization. (2016) Noncommunicable diseases (ncd) country profiles. [En línea]. Disponible: [https://www.who.int/es/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/es/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds))
- [3] W. H. Organization. (2016) Cardiovascular disease. [En línea]. Disponible: http://www.who.int/cardiovascular_diseases/world-heart-day/en/
- [4] T. Hubbner, A.Heitmann and R.Schroeder, “Multivariate short-term heart rate variability: a pre-diagnostic tool for screening heart disease,” *Medical Biological Engineering Computing*, 2011.
- [5] M. P. Tulppo, A. M. Kiviniemi, A. J. Hautala, M. Kallio, T. Seppanen, T. H. Makikallio, and H. V. Huikuri, “Physiological background of the loss of fractal heart rate dynamics,” *Circulation-Hagertown*, vol. 112, no. 3, pp. 314–319, 2005.
- [6] J. Starkova and V.Sovoliev, “Particular work qualities of the complex cardiac system of a human organism,” *Kyiv National Economic University*, 2009.
- [7] M. Mohebbi and H. Ghassemian, “Prediction of paroxysmal atrial fibrillation based on non-linear analysis and spectrum and bispectrum features of the heart rate variability signal,” *Computer methods and programs in biomedicine*, vol. 105, no. 1, pp. 40–49, 2012.
- [8] R. Kemppainen *et al.*, “Ecg parameters in short-term prediction of ventricular arrhythmias,” 2012.
- [9] S. Sahay, A. Wadhwan, and S. Wadhwan, “A survey approach on ecg feature extraction techniques,” *International Journal of Computer Applications*, vol. 120, no. 11, 2015.
- [10] Y. Öztürk, “Fractal dimension as a diagnostic tool for cardiac diseases,” 2019.
- [11] S. Guth and T. P. Sapsis, “Machine learning predictors of extreme events occurring in complex dynamical systems,” *Entropy*, vol. 21, no. 10, p. 925, 2019.
- [12] M. Khalil, A. Adib *et al.*, “An end-to-end multi-level wavelet convolutional neural networks for heart diseases diagnosis,” *Neurocomputing*, vol. 417, pp. 187–201, 2020.
- [13] G. Georgieva-Tsaneva, “Investigation of heart rate variability by statistical methods and detrended fluctuation analysis,” 2019.
- [14] Y. Contoyiannis, F. Diakonos, and M. Kampitakis, “Applying the method of critical fluctuations on human electrocardiograms,” *arXiv preprint arXiv:1908.06408*, 2019.
- [15] D. Makowiec and R. Galaska, “Long-range dependencies in heart rate signals- revisited,” *1st Department of Cardiology, Medical University of Gdańsk, Poland*, 2008.
- [16] ——, “Predictors of congestive heart failure in the elderly: the cardiovascular health study,” *Journal of the American College of Cardiology*, 2000.
- [17] E. Wilkins, L. Wilson, K. Wickramasinghe, P. Bhatnagar, J. Leal, R. Luengo-Fernandez, R. Burns, M. Rayner, and N. Townsend, “European cardiovascular disease statistics 2017,” 2017.

- [18] D. Charles and J.King, "Adoption of electronic health record systems among u.s. non-federal acute care hospitals," *ONC Data Brief No. 9*, 2013.
- [19] T. Pollard, A.Johnson and L.Shen, "Mimic-iii, a freely accessible critical care database," *ONC Data Brief No. 9*, 2013.
- [20] S. D. Ghiassian, "Network medicine: a network-based approach to human diseases." 2015.
- [21] M.-W. Dictionary, "Merriam-webster," *On-line at <http://www.mw.com/home.htm>*, 2002.
- [22] T. Rikitake and K. Hamada, "Earthquake prediction," in *Encyclopedia of Physical Science and Technology (Third Edition)*, third edition ed., R. A. Meyers, Ed. New York: Academic Press, 2003, pp. 743 – 760. [En línea]. Disponible: <http://www.sciencedirect.com/science/article/pii/B0122274105001952>
- [23] A. Carpinteri and O. Borla, "Fracto-emissions as seismic precursors," *Engineering Fracture Mechanics*, vol. 177, pp. 239 – 250, 2017. [En línea]. Disponible: <http://www.sciencedirect.com/science/article/pii/S0013794416306877>
- [24] A. Guyton and J.E.Hall, "*Músculo cardiaco: el corazón como bomba y la función de las válvulas cardíacas*". Elsevier: 2011, Tratado de fisiología médica, 2011,pp.101-115.
- [25] ——, *.Excitación rítmica del corazón*. Elsevier: 2011, Tratado de fisiología médica, 2011,pp.116-122.
- [26] ——, *.El electrocardiograma normal*. Elsevier: 2011, Tratado de fisiología médica, 2011,pp.123-128.
- [27] ——, *Interpretación electrocardiográfica del músculo cardíaco y anomalías coronarias: análisis vectorial*. Elsevier: 2011, Tratado de fisiología médica, 2011,pp.129-142.
- [28] R. Barbieri, E. P. Scilingo, and G. Valenza, *Complexity and nonlinearity in cardiovascular signals*. Springer, 2017.
- [29] J. W. Waks and M. E. Josephson, "Mechanisms of atrial fibrillation–reentry, rotors and reality," *Arrhythmia & electrophysiology review*, vol. 3, no. 2, p. 90, 2014.
- [30] C. Antzelevitch and A. Burashnikov, "Overview of basic mechanisms of cardiac arrhythmia," *Cardiac electrophysiology clinics*, vol. 3, no. 1, pp. 23–45, 2011.
- [31] V. S. Vasudevan, M. A. Simaan, and P. D. Wearden, "Aortic valve ejection fraction for monitoring heart contractility in patients supported with a continuous flow left ventricular assist device," 2019.
- [32] K. Yamaoka, T. Nakagawa, and T. Uno, "Statistical moments in pharmacokinetics," *Journal of pharmacokinetics and biopharmaceutics*, vol. 6, no. 6, pp. 547–558, 1978.
- [33] A. V. Oppenheim, A. S. Willsky, and S. H. Nawab, *Señales y sistemas*. Pearson Educación, 1998.
- [34] P. J. Brockwell, R. A. Davis, and S. E. Fienberg, *Time series: theory and methods: theory and methods*. Springer Science & Business Media, 1991.
- [35] J. Machta, "Entropy, information, and computation," *American Journal of Physics*, vol. 67, no. 12, pp. 1074–1077, 1999.

- [36] J. S. Richman, D. E. Lake, and J. R. Moorman, “Sample entropy,” in *Methods in enzymology*. Elsevier, 2004, vol. 384, pp. 172–184.
- [37] S. M. Pincus, “Approximate entropy as a measure of system complexity.” *Proceedings of the National Academy of Sciences*, vol. 88, no. 6, pp. 2297–2301, 1991.
- [38] J. Feder, *Fractals*. Springer Science & Business Media, 2013.
- [39] (2020) Fractal objects and self-similar processes. [En línea]. Disponible: <http://reylab.bidmc.harvard.edu/tutorial/DFA/node3.html>
- [40] Y. Ashkenazy, P. C. Ivanov, S. Havlin, C. Peng, Y. Yamamoto, A. Goldberger, and H. Stanley, “Decomposition of heartbeat time series: scaling analysis of the sign sequence,” in *Computers in Cardiology 2000. Vol. 27 (Cat. 00CH37163)*. IEEE, 2000, pp. 139–142.
- [41] (2020) Detrended fluctuation analysis (dfa). [En línea]. Disponible: <http://reylab.bidmc.harvard.edu/tutorial/DFA/node5.html>
- [42] C. K. Karmakar, J. Gubbi, A. H. Khandoker, and M. Palaniswami, “Analyzing temporal variability of standard descriptors of poincaré plots,” *Journal of Electrocardiology*, vol. 43, no. 6, pp. 719 – 724, 2010. [En línea]. Disponible: <http://www.sciencedirect.com/science/article/pii/S0022073610003821>
- [43] W. Research, “MorletWavelet,” <https://reference.wolfram.com/language/ref/MorletWavelet.html>, 2010.
- [44] A. N. Akansu and R. A. Haddad, “Chapter 6 - wavelet transform,” in *Multiresolution Signal Decomposition (Second Edition)*, second edition ed., A. N. Akansu and R. A. Haddad, Eds. San Diego: Academic Press, 2001, pp. 391 – 442. [En línea]. Disponible: <http://www.sciencedirect.com/science/article/pii/B9780120471416500069>
- [45] C. Mittermayr, B. Lendl, E. Rosenberg, and M. Grasserbauer, “The application of the wavelet power spectrum to detect and estimate 1/f noise in the presence of analytical signals,” *Analytica Chimica Acta*, vol. 388, no. 3, pp. 303 – 313, 1999. [En línea]. Disponible: <http://www.sciencedirect.com/science/article/pii/S0003267099000835>
- [46] L. Kristoufek, “Fractal markets hypothesis and the global financial crisis: Wavelet power evidence,” *Scientific reports*, vol. 3, no. 1, pp. 1–7, 2013.
- [47] C. Molnar, “Interpretable machine learning,” *Lulu. com*, 2019.
- [48] R. Guerequeta and A. Vallecillo, *Técnicas de diseño de algoritmos*, 2019.
- [49] MIT Laboratory for Computational Physiology. (2015) Python waveform-database (wfdb) package. [En línea]. Disponible: <https://github.com/MIT-LCP/wfdb-python>
- [50] S. J. Orfanidis, *Introduction to Signal Processing*, 1995.

Apéndice

A: Resultados de Análisis

Histogramas y Distribuciones

Histograma de técnicas lineales para Casos Saludables y de ECV

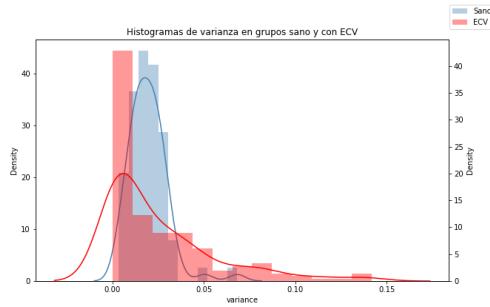


Figura 1: Histograma de varianzas casos sanos y ECV.

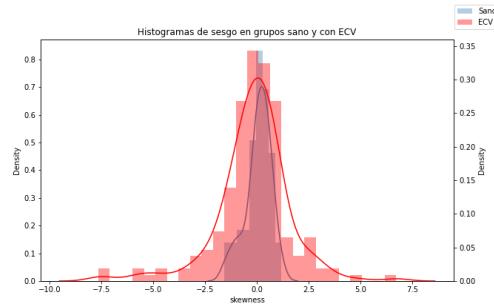


Figura 2: Histograma de Sesgo casos sanos y ECV.

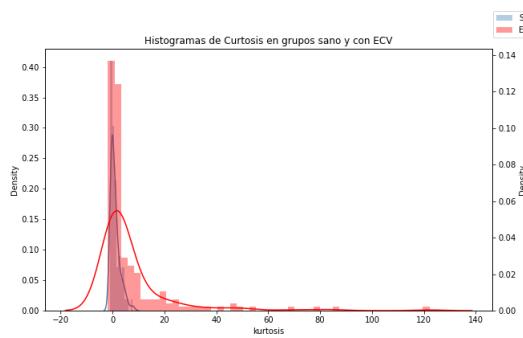


Figura 3: Histograma de Curtosis casos sanos y ECV .

Histogramas de técnicas no lineales para Casos sanos y ECV.

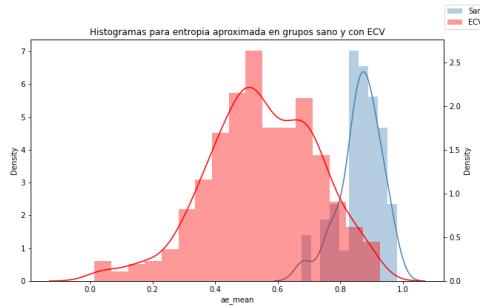


Figura 4: Histograma de entropía aproximada para casos sanos y con ECV

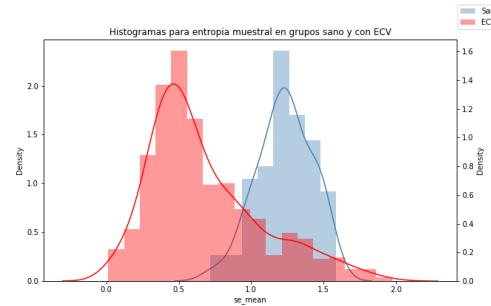


Figura 5: Histograma de entropía muestral para casos sanos y con ECV

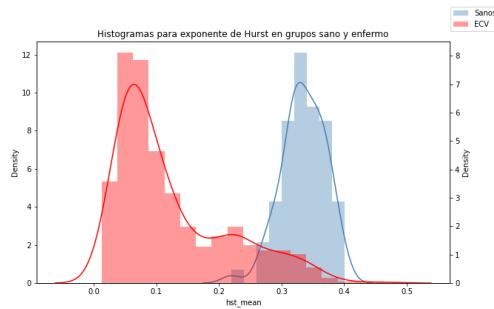


Figura 6: Histograma de exponente de Hurst para casos sanos y con ECV

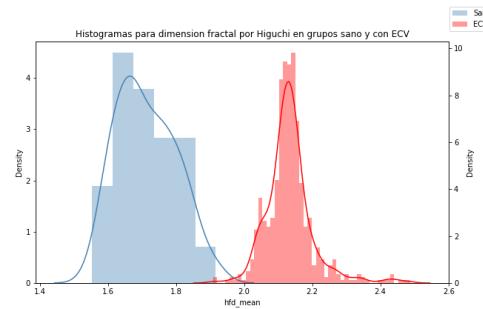


Figura 7: Histograma de dimensión fractal por Higuchi para casos sanos y con ECV

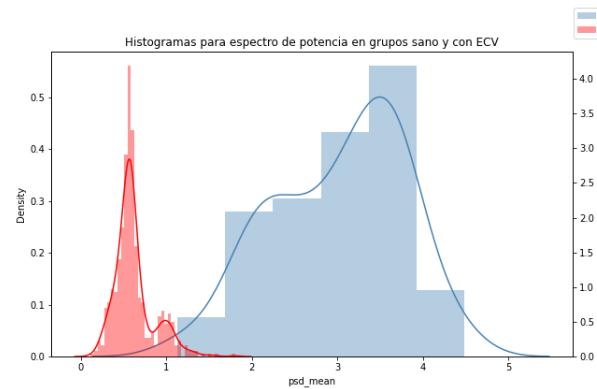


Figura 8: Histograma del espectro de potencia para casos sanos y con ECV

Histogramas de técnicas no lineales para Casos de FA,ICC e IM

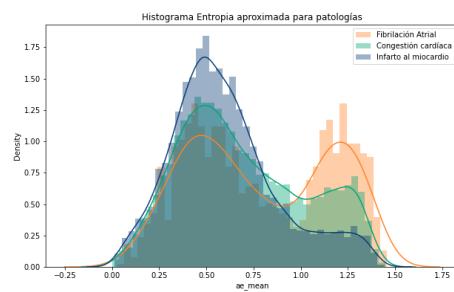


Figura 9: Histograma de entropía aproximada para las 3 patologías

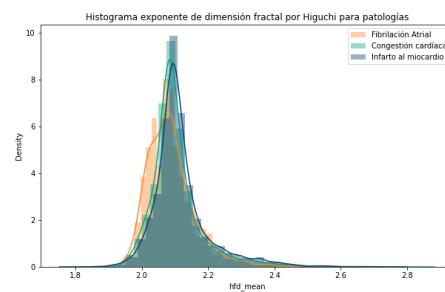


Figura 10: Histograma de dimensión fractal por Higuchi para las 3 patologías

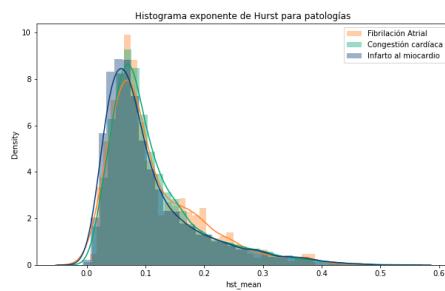


Figura 11: Histograma de exponente de Hurst para las 3 patologías

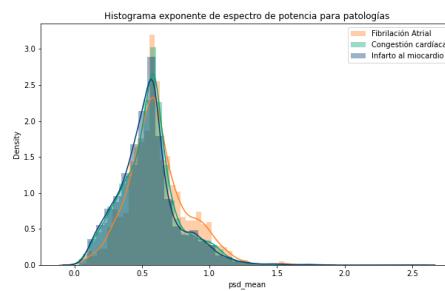


Figura 12: Histograma de espectro de potencia para las 3 patologías

Gráficas de dispersión de técnicas lineales para casos sanos vs ECV

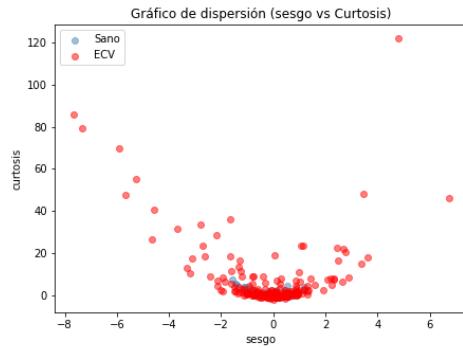


Figura 13: Gráfica de dispersión sesgo vs Curtosis

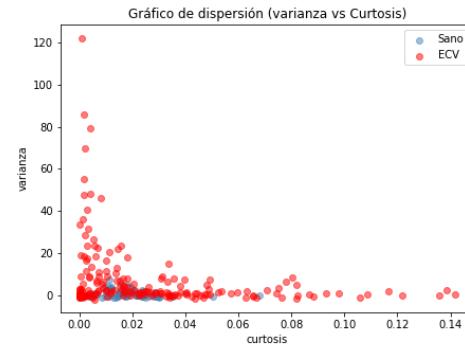


Figura 14: Gráfico de dispersión varianza vs Curtosis

Gráficas de dispersión de técnicas no lineales para casos sanos vs ECV

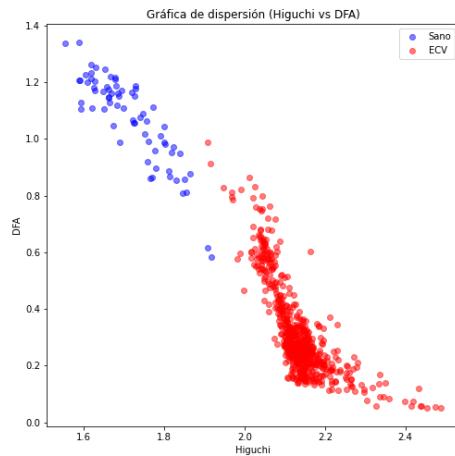


Figura 15: Gráfica de dispersión Higuchi vs DFA

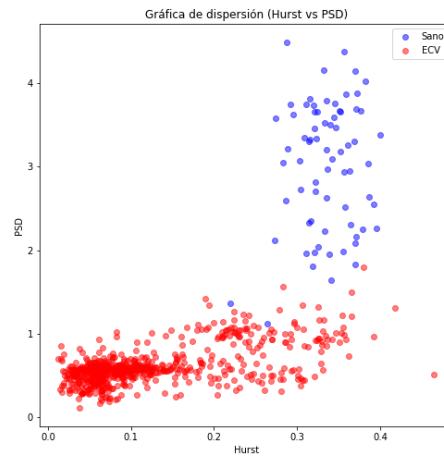


Figura 16: Gráfico de dispersión Hurst vs PSD

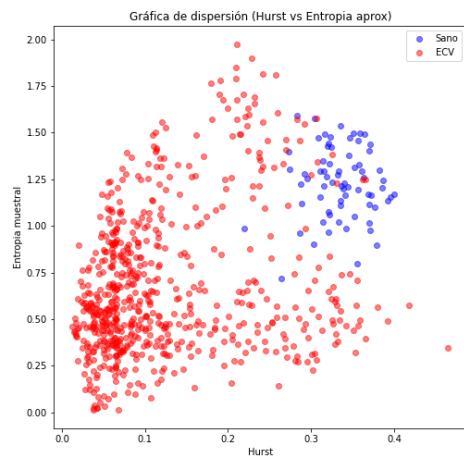


Figura 17: Gráfica de dispersión Hurst vs entropía muestral

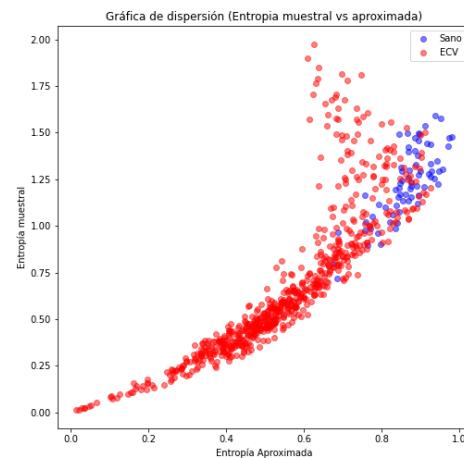


Figura 18: Gráfico de dispersión Entropía muestral vs entropía aproximada

Gráficas de dispersión de técnicas no lineales para casos FA,ICC e IM

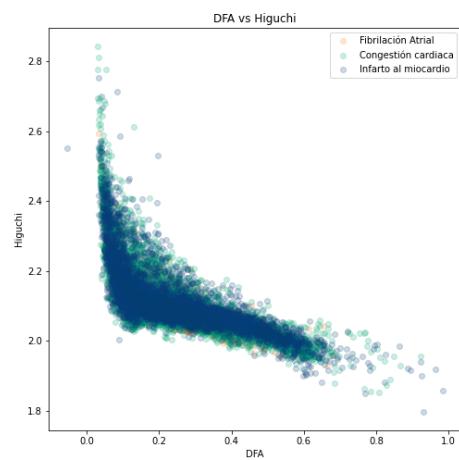


Figura 19: Gráfica de dispersión DFA vs Higuchi

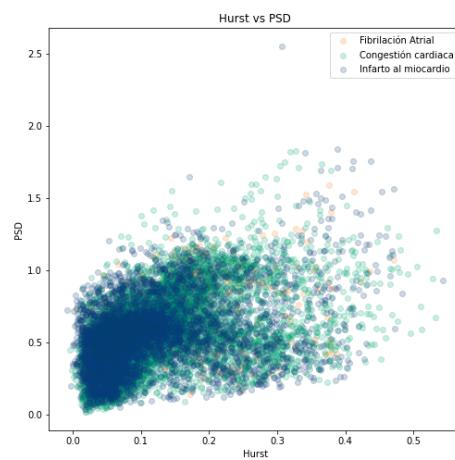


Figura 20: Gráfico de dispersión exponente de Hurst vs espectro de potencia

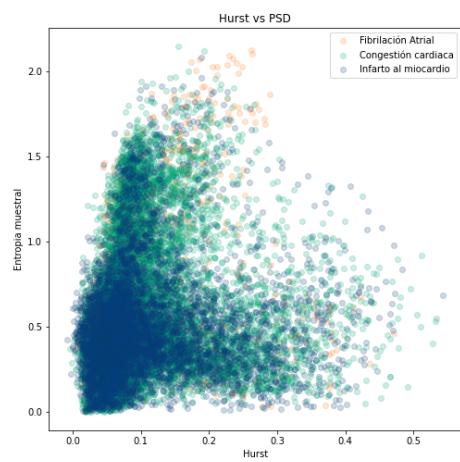


Figura 21: Gráfica de dispersión exponente de Hurst vs entropía muestral

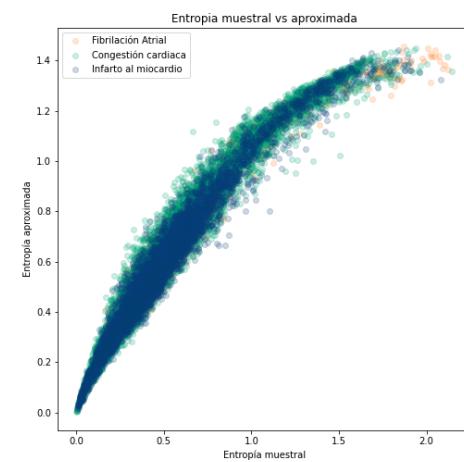


Figura 22: Gráfico de dispersión entropía muestral vs aproximada

B: Resultados temporales

Resultados lineales de conjuntos de Fibrilación Atrial



Figura 23: Metricas lineales de casos de FA

Resultados lineales de conjuntos de Insuficiencia Cardíaca Congestiva



Figura 24: Metricas lineales de casos de ICC

Resultados lineales de conjuntos de Casos Saludables



Figura 25: Metricas lineales de casos de Control

Resultados no-lineales de conjuntos de Fibrilación Atrial

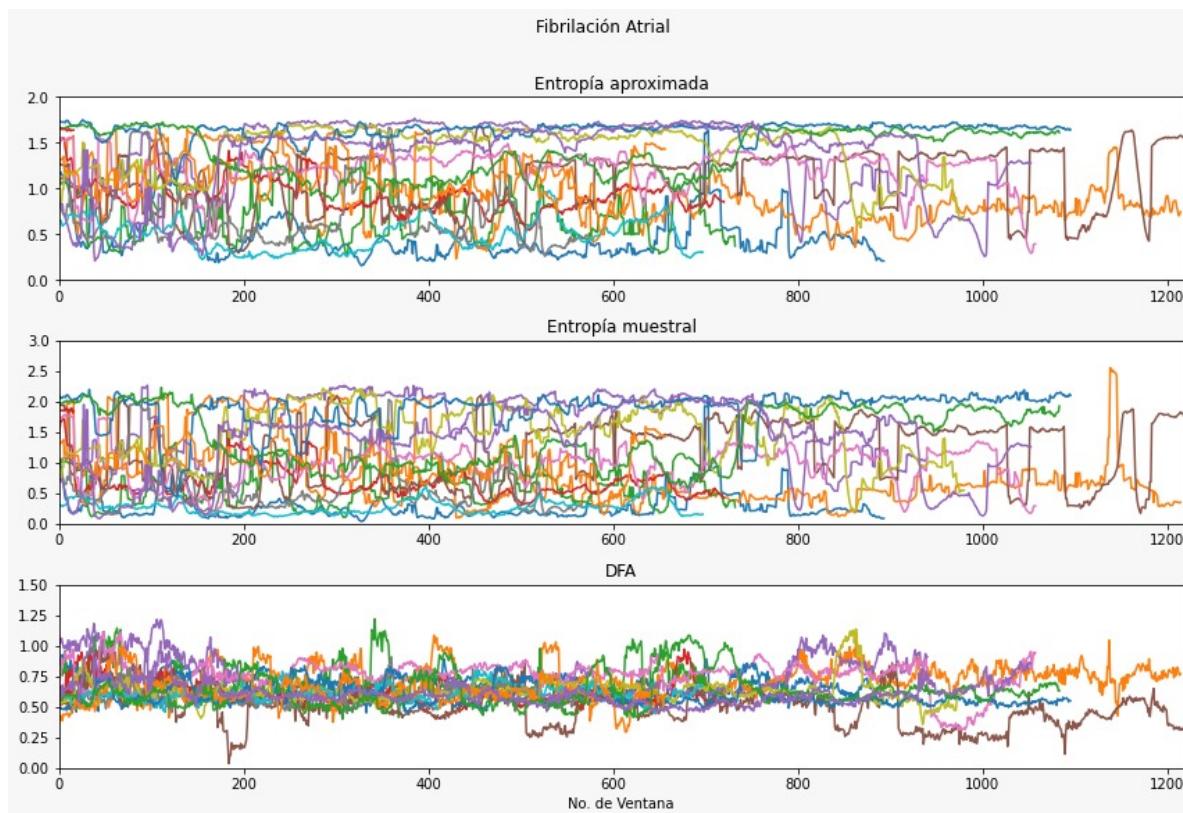


Figura 26: Metricas no-lineales de casos de FA

Resultados no-lineales de conjuntos de Insuficiencia Cardíaca Congestiva

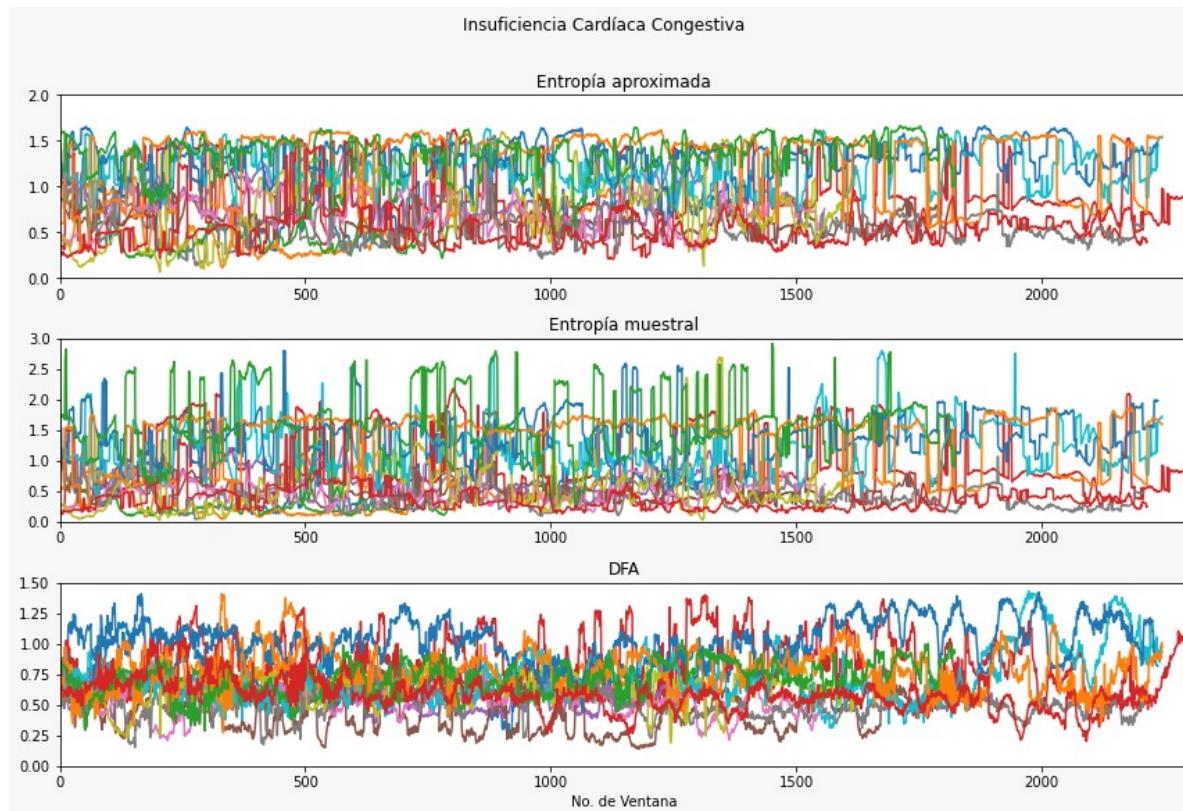


Figura 27: Metricas no-lineales de casos de ICC

Resultados no-lineales de conjuntos de Casos Saludables

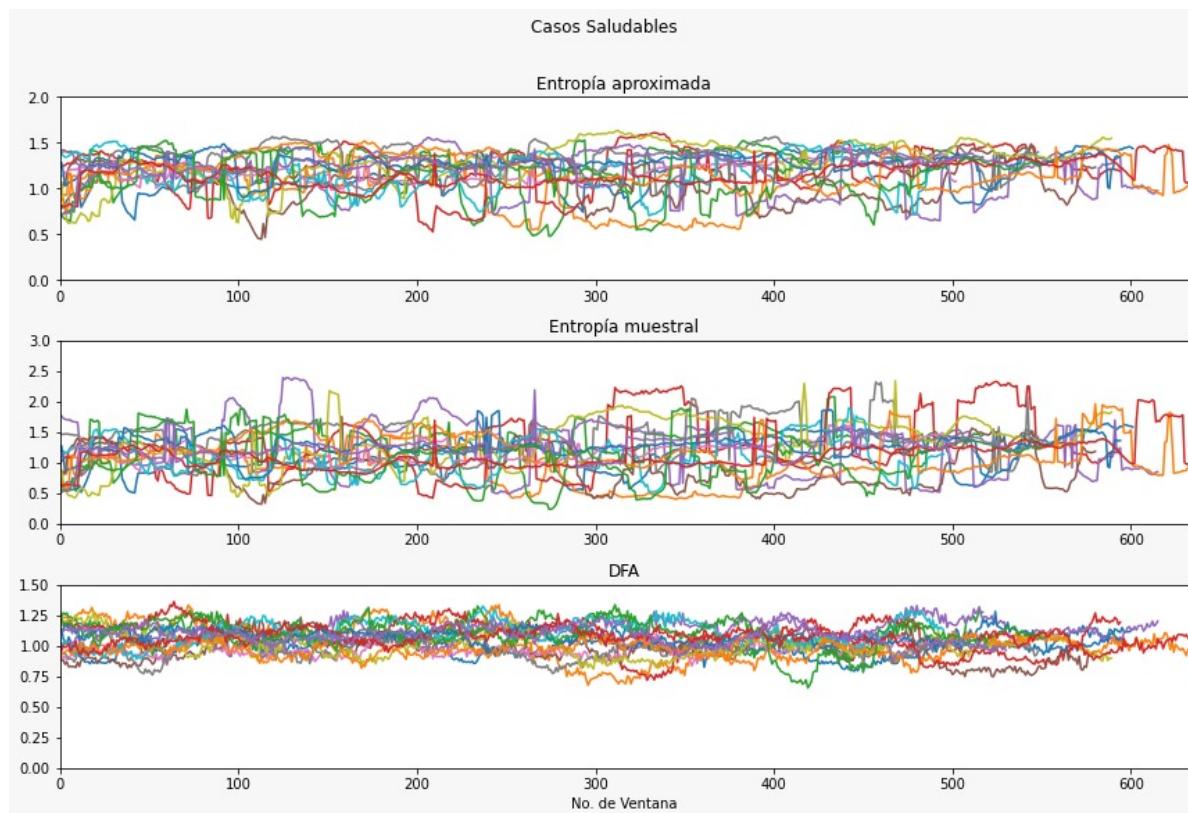


Figura 28: Metricas no-lineales de casos de Control

Repository del Proyecto

Para mayor referencia ó consulta del código utilizado se ha creado este proyecto en la plataforma GitHub. A continuación se presenta un enlace al repositorio que aloja este proyecto.

- https://github.com/AlanPalomino/TTB_2020_2_13

Repositorios utilizados en el Proyecto

Así mismo se menciona a continuación los repositorios que fueron consultados e incluidos en el código de este proyecto. Su valiosa aportación es inmensamente agradecida.

- <https://github.com/raphaelvallat/entropy>
- <https://github.com/PGomes92/pyhrv>